GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 21, 2003, 03:33:17 ; Search time 1431 Seconds
(without alignments)
1459.971 Million cell updates/sec

Title: US-09-689-430-1_COPY_150_278
Perfect score: 129
Sequence: 1 ctctttctaagtaaacagta.....gccatcagcgcatgcggatc 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	59.2	45.9	479	10	AV657611	AV657611 AV657611
	2	46	35.7	451	10	AV685661	AV685661 AV685661
	3	40.8	31.6	421	10	AV654002	AV654002 AV654002
	4	32.8	25.4	927	14	BQ642439	BQ642439 AGENCOURT
c	5	32.6	25.3	855	12	BF144380	BF144380 601787403
c	6	32.2	25.0	466	9	AA616160	AA616160 vo92g05.r
c	7	31	24.0	915	13	BI561624	BI561624 603256037
	8	30.8	23.9	928	13	BI868197	BI868197 603392138
c	9	30.6	23.7	213	10	AW445510	AW445510 81721 MAR
	10	30.6	23.7	348	12	BF835894	BF835894 RC1-HT097
	11	30.6	23.7	377	12	BF835886	BF835886 RC1-HT097
c	12	30.6	23.7	394	13	BM430717	BM430717 1Duo03G07
c	13	30.6	23.7	458	13	BM429757	BM429757 1Duo20H8.
c	14	30.6	23.7	552	14	BQ355187	BQ355187 MR2-HT116
c	15	30.6	23.7	567	10	AV596621	AV596621 AV596621
c	16	30.6	23.7	568	10	AV616893	AV616893 AV616893
c	17	30.6	23.7	750	10	BE296329	BE296329 601174343
c	18	30.6	23.7	765	10	BE513735	BE513735 601315227
c	19	30.6	23.7	915	14	BQ648846	BQ648846 AGENCOURT
c	20	30.6	23.7	939	14	BQ709796	BQ709796 AGENCOURT
c	21	30.6	23.7	951	14	BQ709274	BQ709274 AGENCOURT
c	22	30.6	23.7	955	14	BQ706617	BQ706617 AGENCOURT
	23	30.6	23.7	1245	14	BQ066190	BQ066190 AGENCOURT
	24	30.4	23.6	899	13	BM014482	BM014482 603640258
	25	30.2	23.4	678	9	AL632211	AL632211 AL632211
	26	30	23.3	599	10	AW953618	AW953618 EST365688
c	27	30	23.3	1063	12	BE871724	BE871724 601449587
	28	29.8	23.1	568	17	AF019127	AF019127 AF019127
	29	29.6	22.9	519	17	AZ902709	AZ902709 RPCI-24-1
	30	29.6	22.9	874	13	BM043559	BM043559 603619992
	31	29.6	22.9	888	13	BM011355	BM011355 603635542
	32	29.6	22.9	1060	13	BM470623	BM470623 AGENCOURT
c	33	29.2	22.6	251	14	F05149	F05149 HSC02E071 n
	34	29.2	22.6	338	12	BF835885	BF835885 RC1-HT097
c	35	29.2	22.6	462	13	BG956025	BG956025 PM2-CT080
	36	29.2	22.6	503	17	AQ853104	AQ853104 LMAJFV1_1
c	37	29.2	22.6	842	10	BE560380	BE560380 601346682
c	38	29	22.5	462	14	BM967767	BM967767 LM24HW011
c	39	29	22.5	500	10	AW502350	AW502350 UI-HF-BR0
	40	29	22.5	743	12	BF862037	BF862037 963033A07
	41	28.8	22.3	431	14	R89172	R89172 yq02b08.r1
c	42	28.8	22.3	441	10	AV636155	AV636155 AV636155
c	43	28.8	22.3	547	10	AV619091	AV619091 AV619091
	44	28.8	22.3	563	14	BQ814899	BQ814899 1030046E0
c	45	28.8	22.3	663	14	BQ824933	BQ824933 1030122F0

ALIGNMENTS

RESULT 1
AV657611
LOCUS AV657611 479 bp mRNA linear EST 16-JAN-2002
DEFINITION AV657611 GLC Homo sapiens cDNA clone GLCFDG01 3', mRNA sequence.
ACCESSION AV657611
VERSION AV657611.1 GI:9878625
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 479)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

```
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCFDG01"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      132 a    104 c    116 g    127 t
ORIGIN

  Query Match             45.9%;  Score 59.2;  DB 10;  Length 479;
  Best Local Similarity   72.6%;  Pred. No. 5.1e-09;
  Matches   90;  Conservative    0;  Mismatches   33;  Indels    1;  Gaps    1;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCT-CGGCAACGGCCTGGTCTGT 61
           || ||| || ||||| ||| |||||||||||||||||| | |||||||| | |||| |
Db     263 CTGTCTGAGGAAACAATACTGGAACCTTTACCCCGTTGATATGGCAACGGGCATGTCTTT 322

Qy      62 GCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCA 121
           | ||||||||||||||| ||   | | | ||  ||  | |||  ||     |||||| |||
Db     323 GTCAAGTGTTTGCTGATGCGTACACTAGTGAATGCAGATTGAACACGTCTCATCAGAGCA 382

Qy     122 TGCG 125
           ||||
Db     383 TGCG 386


RESULT 2
AV685661
LOCUS       AV685661                451 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV685661 GKC Homo sapiens cDNA clone GKCCGH01 5', mRNA sequence.
ACCESSION   AV685661
VERSION     AV685661.1  GI:10287524
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GKCCGH01"
                     /clone_lib="GKC"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       88 a    126 c    124 g    113 t
ORIGIN

  Query Match             35.7%;  Score 46;  DB 10;  Length 451;
  Best Local Similarity   63.6%;  Pred. No. 0.00012;
  Matches   70;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      16 CAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTTGCT 75
           | |||  | ||| | ||| |||||| ||||||| ||          |   ||||  ||||
Db       1 CTGTAAGTCAACATATACGCCGTTGATCGGCAATGGTAACTGTGCAGAATAGTGGATGCT 60

Qy      76 GACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCG 125
           ||||||  | | |||||    |||| |||||| |||||  |   | ||||
Db      61 GACGCATACGCTACTGGTGCTGGCTGGGCCATGGGCCAAGAATCCGTGCG 110


RESULT 3
AV654002
LOCUS       AV654002                421 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV654002 GLC Homo sapiens cDNA clone GLCDQH06 3', mRNA sequence.
ACCESSION   AV654002
VERSION     AV654002.1  GI:9875016
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCDQH06"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      112 a     99 c     82 g    128 t
ORIGIN

  Query Match             31.6%;  Score 40.8;  DB 10;  Length 421;
  Best Local Similarity   77.5%;  Pred. No. 0.0063;
  Matches   62;  Conservative    0;  Mismatches   17;  Indels    1;  Gaps    1;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTT-TACCCCGTTGCTCGGCAACGGCCTGGTCTGT 61
           ||| ||  ||||||| |   | || ||| |||||||||  || |||||| |  ||||||
Db     342 CTTACTGCGTAAACAATGTCTAAATCTTATACCCCGTTGACCGTCAACGGTCATGTCTGT 401

Qy      62 GCCAAGTGTTTGCTGACGCA 81
           ||||| || ||||||||||
Db     402 GCCAACTGATTGCTGACGCA 421


RESULT 4
BQ642439
LOCUS       BQ642439                927 bp    mRNA    linear   EST 15-JUL-2002
```

```
DEFINITION  AGENCOURT_8286037 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:6292602
            5', mRNA sequence.
ACCESSION   BQ642439
VERSION     BQ642439.1  GI:21766611
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2494  row: e  column: 19
            High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6292602"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      214 a    266 c    244 g    202 t      1 others
ORIGIN

  Query Match             25.4%;  Score 32.8;  DB 14;  Length 927;
  Best Local Similarity   59.8%;  Pred. No. 4.3;
  Matches   55;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      21 CATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTTGCTGACGC 80
           ||   |||||||| || | |   ||    | |||  ||   |||  | | || | | |
Db     757 CAAACACCTTTACACCCTAGAAGGGGGGGGACCTCATCAAGGCCCTGGGCTTTCAGCCCT 816

Qy      81 AACCCCCACTGGCTGGGGCTTGGCCATAGGCC 112
           |||| | |||||  |||||  |||||   | ||
Db     817 AACCGCTACTGGTTGGGGGCTGGCCCAGGCCC 848


RESULT 5
BF144380/c
LOCUS       BF144380                855 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  601787403F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4015105 5',
            mRNA sequence.
ACCESSION   BF144380
VERSION     BF144380.1  GI:10983420
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9260  row: f  column: 02
            High quality sequence stop: 693.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4015105"
                     /clone_lib="NCI_CGAP_Lu30"
                     /tissue_type="tumor, metastatic to mammary"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
                     dT.   Library constructed by Life Technologies.
                     Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      188 a    211 c    279 g    177 t
ORIGIN

  Query Match             25.3%;  Score 32.6;  DB 12;  Length 855;
  Best Local Similarity   69.8%;  Pred. No. 4.8;
  Matches   44;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       1 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 60
           |||| |||||||  |||| || || | ||||   ||| |  || ||| | ||   |||||
Db     281 CTCTCTCTAAGTCCACAGCACCTGGATCTTTGAGCCGGTCTTCAGCAGCTGCTGCGTCTG 222

Qy      61 TGC 63
           |||
Db     221 TGC 219


RESULT 6
AA616160/c
LOCUS       AA616160                466 bp    mRNA    linear   EST 07-OCT-1997
DEFINITION  vo92g05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1066616 5' similar to gb:M64716 40S RIBOSOMAL PROTEIN
            S25 (HUMAN);, mRNA sequence.
ACCESSION   AA616160
VERSION     AA616160.1  GI:2503365
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:588976
            Seq primer: -28m13 rev2 ET from Amersham
```

```
          High quality sequence stop: 107.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1066616"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG], digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT      143 a     99 c    129 g     95 t
ORIGIN

  Query Match             25.0%;  Score 32.2;  DB 9;  Length 466;
  Best Local Similarity   54.7%;  Pred. No. 4.8;
  Matches   64;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy       2 TCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGT 61
           | || | || |   || |   | ||||||| |  ||    |  ||||   | |||   ||
Db     462 TTTTCCAAAATTCCCACTTGTTAAACCTTTTCAACGAATTTTCGCAAGCACTTGGAGCGT 403

Qy      62 GCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGC 118
             ||||   |||| ||   |       | |  |||||||||  ||| | ||  | ||||
Db     402 CCCAACCCTTTGGTGTTCCGGGTGTAAATAACTGGGGCTCTGCCCTTGGAAACCAGC 346


RESULT 7
BI561624/c
LOCUS       BI561624                915 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603256037F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5298512 5',
            mRNA sequence.
ACCESSION   BI561624
VERSION     BI561624.1  GI:15448938
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11755  row: i  column: 09
            High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1. .915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5298512"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      248 a    192 c    257 g    218 t
ORIGIN

  Query Match             24.0%;  Score 31;  DB 13;  Length 915;
  Best Local Similarity   52.8%;  Pred. No. 17;
  Matches   67;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           | |  |||||  ||| | || |  | |  |  ||| || ||  |   | |  ||    ||
Db     880 CCTGGTAAGTCCACATTGCAAGCCCATGGAATCCGATGGTCCCCTTGGCCAGGGATAATG 821

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
             ||  | ||  || |   |   | |  |||  |||| ||| ||| |||||      |||
Db     820 AGAAACGCTTATTGCCAAGATGTCAATAGGCATGGGCATGGGCATTGGCCACGCTTCCAT 761

Qy     123 GCGGATC 129
           | | ||
Db     760 CCAGGTC 754


RESULT 8
BI868197
LOCUS       BI868197                928 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  603392138F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5402265 5',
            mRNA sequence.
ACCESSION   BI868197
VERSION     BI868197.1  GI:16041870
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12025  row: l  column: 10
            High quality sequence stop: 840.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5402265"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      222 a    272 c    237 g    197 t
ORIGIN

  Query Match             23.9%;  Score 30.8;  DB 13;  Length 928;
  Best Local Similarity   57.1%;  Pred. No. 20;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;
```

```
Qy      21 CATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTTGCTGACGC 80
           ||   |||||||| |  | | | |     | | |  ||   |||  |||  | | | |
Db     738 CAAACACCTTTACACGCTAGATGGTGGGGGACATCATCAACGCCCTGTGGCTTCAGCCCT 797

Qy      81 AACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGC 118
           |||| | |||||||| |   || |||  | ||| || |
Db     798 AACCGCTACTGGCTGTGTGCTGCCCAGGGCCCAGCATC 835


RESULT 9
AW445510/c
LOCUS       AW445510                213 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  81721 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW445510
VERSION     AW445510.1  GI:6987272
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 213)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 43  row: A  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .213
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT       46 a     60 c     64 g     43 t
ORIGIN

  Query Match            23.7%;  Score 30.6;  DB 10;  Length 213;
  Best Local Similarity   60.0%;  Pred. No. 11;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      11 GTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGT 70
           || | ||| ||| | ||||   || || ||| | | || |||| |      || ||| |
Db     106 GTGACCAGCACAAGCACCTCGGCCACGATGCCCAGGAAGGGCCAGAGGGCAGCAAAGCGG 47

Qy      71 TTGCTGACGCAACCCCCACTGGCTG 95
           ||||  ||||    |    ||||||
Db      46 TTGCGCACGCGCAACGTGATGGCTG 22


RESULT 10
BF835894
LOCUS       BF835894                348 bp    mRNA    linear   EST 13-JAN-2001
DEFINITION  RC1-HT0975-161100-021-f05 HT0975 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF835894
VERSION     BF835894.1  GI:12187336
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-HT0975-
            161100-021-f05&t3=2000-11-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 348.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0975"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       65 a    102 c    125 g     56 t
ORIGIN

  Query Match            23.7%;  Score 30.6;  DB 12;  Length 348;
  Best Local Similarity   53.8%;  Pred. No. 14;
  Matches   63;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      13 AAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTT 72
           ||| |  |  || |||    | |.|  | | | |   | |||| |    ||| || |  |
Db     162 AAAGATGAGCTGCACCACGTCACTGAGGTTGGCCGGGGCCCTGCTGAAAGCCTAGGGGAT 221

Qy      73 GCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCGGATC 129
           |  || |  |  |||    |||||||  |||||||  || |||     |||| ||
Db     222 GGAGAAGTGGCAGCCAGGCCCTGGGGCGCCGCCATAGCACAGCAGGCTCCGCGGGTC 278


RESULT 11
BF835886
LOCUS       BF835886                377 bp    mRNA    linear   EST 13-JAN-2001
DEFINITION  RC1-HT0975-161100-021-a09 HT0975 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF835886
VERSION     BF835886.1  GI:12187321
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=RC1&t2=RC1-HT0975-
            161100-021-a09&t3=2000-11-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 377.
FEATURES                Location/Qualifiers
     source             1. .377
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="HT0975"
                        /dev_stage="Adult"
                        /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                        Site_2: SmaI; A mini-library was made by cloning products
                        derived from  ORESTES PCR (U.S. Letters Patent application
                        No. 196,716 -  Ludwig Institute for Cancer Research)
                        profiles into  the pUC 18 vector. Reverse transcription of
                        tissue mRNA and cDNA amplification were performed under
                        low stringency conditions."
BASE COUNT       68 a    108 c    136 g     65 t
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 12;  Length 377;
  Best Local Similarity   53.8%;  Pred. No. 15;
  Matches   63;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      13 AAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTT 72
           ||| |  |  || |||    | | |  | | | |   | |||| |    ||| || |  |
Db     178 AAAGATGAGCTGCACCACGTCACTGAGGTTGGCCGGGGCCCTGCTGAAAGCCTAGGGGAT 237

Qy      73 GCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCGGATC 129
           |  || |   |  |||    |||||||   |||||||  || |||     |||| ||
Db     238 GGAGAAGTGGCAGCCAGGCCCTGGGGCGCCGCCATAGCACAGCAGGCTCCGCGGGTC 294


RESULT 12
BM430717/c
LOCUS       BM430717                394 bp    mRNA    linear   EST 31-JAN-2002
DEFINITION  1Duo03G07 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
            sequence.
ACCESSION   BM430717
VERSION     BM430717.1  GI:18452439
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
            ,P.M.K. and Moore,S.S.
  TITLE     Gene Expression Profiling of the Bovine Gastrointestinal Tract
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Stephen Moore
            . Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoore@afns.ualberta.ca
            Insert Length: 394   Std Error: 0.00
            POLYA=No.
FEATURES                Location/Qualifiers
     source             1. .394
                        /organism="Bos taurus"
                        /db_xref="taxon:9913"
                        /clone_lib="Bos taurus Duodenum #1 library"
                        /tissue_type="Smooth muscle"
                        /cell_type="Simple columnar epithelial"
                        /dev_stage="Young adult"
                        /lab_host="XL1-BlueMRF'strain"
                        /note="Organ: Intestine/duodenum; Vector: Uni-2ZAPXR;
                        Site_1: EcoRI; Site_2: Xho I"
BASE COUNT       92 a    120 c    111 g     71 t
ORIGIN

  Query Match             23.7%;  Score 30.6;  DB 13;  Length 394;
  Best Local Similarity   60.0%;  Pred. No. 15;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      11 GTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGT 70
           || | ||| ||| | |||| || || ||| | | || |||| |       || ||| |
Db     291 GTGACCAGCACAAGCACCTCGGCCACGATGCCCAGGAAGGGCCAGAGGGCAGCAAAGCGG 232

Qy      71 TTGCTGACGCAACCCCCACTGGCTG 95
           ||||  ||||    |    ||||||
Db     231 TTGCGCACGCGCAACGTGATGGCTG 207


RESULT 13
BM429757/c
LOCUS       BM429757                458 bp    mRNA    linear   EST 31-JAN-2002
DEFINITION  1Duo20H8.ab1 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
            sequence.
ACCESSION   BM429757
VERSION     BM429757.1  GI:18451479
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
            ,P.M.K. and Moore,S.S.
  TITLE     Gene Expression Profiling of the Bovine Gastrointestinal Tract
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Stephen Moore
            . Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoore@afns.ualberta.ca
            Insert Length: 458   Std Error: 0.00
            POLYA=No.
FEATURES                Location/Qualifiers
     source             1. .458
                        /organism="Bos taurus"
                        /db_xref="taxon:9913"
                        /clone_lib="Bos taurus Duodenum #1 library"
                        /tissue_type="Smooth muscle"
                        /cell_type="Simple columnar epithelial"
                        /dev_stage="Young adult"
```

```
                     /lab_host="XL1-BlueMRF'strain"
                     /note="Organ: Intestine/duodenum; Vector: Uni-2ZAPXR;
                     Site_1: EcoRI; Site_2: Xho I"
BASE COUNT       119 a    139 c    126 g     74 t
ORIGIN

  Query Match           23.7%;  Score 30.6;  DB 13;  Length 458;
  Best Local Similarity   60.0%;  Pred. No. 16;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      11 GTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGT 70
           || | ||| ||| | ||||   || || ||| | | || |||| |      || ||| |
Db     236 GTGACCAGCACAAGCACCTCGGCCACGATGCCCAGGAAGGGCCAGAGGGCAGCAAAGCGG 177

Qy      71 TTGCTGACGCAACCCCCACTGGCTG 95
           ||||  ||||    |    ||||||
Db     176 TTGCGCACGCGCAACGTGATGGCTG 152


RESULT 14
BQ355187/c
LOCUS       BQ355187                552 bp    mRNA    linear   EST 20-MAY-2002
DEFINITION  MR2-HT1160-240101-008-d05 HT1160 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ355187
VERSION     BQ355187.1  GI:21019243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-HT1160-
            240101-008-d05&t3=2001-01-24&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 20
            High quality sequence stop: 59.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT1160"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       102 a    178 c    173 g     99 t
ORIGIN

  Query Match           23.7%;  Score 30.6;  DB 14;  Length 552;
  Best Local Similarity   53.8%;  Pred. No. 18;
  Matches   63;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      13 AAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGTTT 72
           ||| |  |  || |||    | | |  | | | |   | |||| |    ||| || | | |
Db     522 AAAGATGAGCTGCACCACGTCACTGAGGTTAGCCGGGGCCCTGCTGAAAGCCTAGGGGAT 463

Qy      73 GCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCGGATC 129
           |  || |   | |||    |||||||   |||||||  || |||     |||| ||
Db     462 GGAGAAGTGGCAGCCAGGCCCTGGGGCGCCGCCATAGCACAGCAGGCTCCGCGGGTC 406


RESULT 15
AV596621/c
LOCUS       AV596621                567 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  AV596621 Bos taurus cartilage fetus Bos taurus cDNA clone
            E1CA026G04 5', mRNA sequence.
ACCESSION   AV596621
VERSION     AV596621.1  GI:9713705
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1CA026G04"
                     /clone_lib="Bos taurus cartilage fetus"
                     /tissue_type="cartilage"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: Sal1; Site_2: Not1; Poly A
                     was deleted from a Not1 site"
BASE COUNT       129 a    175 c    166 g     96 t      1 others
ORIGIN

  Query Match           23.7%;  Score 30.6;  DB 10;  Length 567;
  Best Local Similarity   60.0%;  Pred. No. 18;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      11 GTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGCCAAGTGT 70
           || | ||| ||| | ||||   || || ||| | | || |||| |      || ||| |
Db     487 GTGACCAGCACAAGCACCTCGGCCACGATGCCCAGGAAGGGCCAGAGGGCAGCAAAGCGG 428

Qy      71 TTGCTGACGCAACCCCCACTGGCTG 95
           ||||  ||||    |    ||||||
Db     427 TTGCGCACGCGCAACGTGATGGCTG 403


Search completed: May 21, 2003, 04:37:33
Job time : 1435 secs
```

```
  JOURNAL   Arch. Virol. 143 (12), 2313-2326 (1998)
  MEDLINE   99129050
REFERENCE   2  (bases 1 to 3221)
  AUTHORS   Mishiro,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1998) Shunji Mishiro, Toshiba General Hospital,
            Dept. Medical Sciences; 6-3-22 Higashi Oh-i, Shinagawa, Tokyo 140,
            Japan (Tel:03-3764-8981, Fax:03-3764-8992)
FEATURES             Location/Qualifiers
     source          1. .3221
                     /organism="Hepatitis B virus"
                     /isolate="11D11HCC"
                     /db_xref="taxon:10407"
                     /note="genotype A-isolated from the circulation of a
                     Japanese patient with hepatocellular carcinoma"
     gene            join(2307. .3221,1. .1623)
                     /gene="P"
     CDS             join(2307. .3221,1. .1623)
                     /gene="P"
                     /codon_start=1
                     /product="pol protein"
                     /protein_id="BAA32871.1"
                     /db_xref="GI:3551317"
                     /translation="MPLSYQHFRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPNIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELQHGRLVIKTSQRHGDESFCSQPSGILSRSSVGPCIRS
                     QLKQSRLGLQPHQGPLASSQPGRSGSIRARVHPSTRRYFGVEPSGSGHIDHSVNNSSS
                     CLHQSAVRKAAYSHLSTSKRQSSSGHAVEFHCLPPNSAGSQSQGSVSSCWWLQFRNSK
                     PCSEYCLSHLVNLREDWGPCDEHGEHHIRIPRTPARVTGGVFLVDKNPHNTAESRLVV
                     DFSQFSRGITRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLL
                     IGSSGLSRYVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSH
                     PIVLGFRKIPMGVGLSPFLLAQFTSAICSVVRRAFPHCLAFSYMDDVVLGAKSVQHRE
                     ALYTAVTNFLLSLGIHLNPNKTKRWGYSLNFMGYIIGSWGTLPQDHIVQKIKHCFRKL
                     PVNRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQ
                     YMNLYPVARQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLPIHTAELLAACFARS
                     RSGAKLIGTDNSVVLSRKYTSFPWLLGCAANWILRGTSFVYVPSALNPADDPSRGRLG
                     LYRPLLRLPFQPTTGRTSLYAVSPSVPSHLPVRVHFASPLHVAWRPP"
     gene            join(2854. .3221,1. .835)
                     /gene="S"
     CDS             join(2854. .3221,1. .835)
                     /gene="S"
                     /codon_start=1
                     /product="envelope protein"
                     /protein_id="BAA32872.1"
                     /db_xref="GI:3551318"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNPDWD
                     FNPIKDHWPAANQVGVGAFGPGFTPPHGGILGWSPQAQGILTTVSTIPPPASTNRQSG
                     RQPTPISPPLRDSHPQAMQWNSTAFHQTLQDPRVRGLYLPAGGSSSGTVNPAPNIASH
                     ISSISARTGDPVTNMENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     SPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGSTTTSTGPCKTCTTPAQGNSMFPSCCCTKPTDGNCTCIPIPSSWAFAK
                     FLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVRPFIPLLPI
                     FFCLWVYI"
     gene            1374. .1838
                     /gene="X"
     CDS             1374. .1838
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="BAA32869.1"
                     /db_xref="GI:3551315"
                     /translation="MAARLCCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQSLPKVLHKRTLGLPA
                     MSTTDLEAYFKDCVFKDWEELGEEIRLMIFVLGGCRHKLVCAPAPCNFFTSA"
     gene            1814. .2458
                     /gene="C"
     CDS             1814. .2458
                     /gene="C"
                     /codon_start=1
                     /product="precore/core protein"
                     /protein_id="BAA32870.1"
                     /db_xref="GI:3551316"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEL
                     LSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWVELMTLATWVG
                     NNLQDPASRDLVVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPA
                     YRPPNAPILSTLPETTVVRRRDRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC"
BASE COUNT      741 a    869 c    707 g    904 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3221;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1116 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1175

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1176 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1235

Qy     123 GCG 125
           |||
Db    1236 GCG 1238


Search completed: May 21, 2003, 04:13:28
Job time : 1028 secs
```

```
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAL68821.1"
                     /db_xref="GI:18389987"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLGLPA
                     MSTTDLEAYFKDCVFKDWEELGEEIRLKVFVLGGCRHKLVCAPAPCNFFTSA"
     gene            1814. .2152
                     /gene="C"
     CDS             1814. .2152
                     /gene="C"
                     /note="pre-C, C"
                     /codon_start=1
                     /product="truncated core antigen"
                     /protein_id="AAL68822.1"
                     /db_xref="GI:18389988"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEL
                     LSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVG
                     NNLEDPHQGI"
     variation       2136^2137
                     /gene="C"
                     /replace="a"
BASE COUNT      746 a    864 c    707 g    903 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3220;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1116 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1175

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1176 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1235

Qy     123 GCG 125
          |||
Db    1236 GCG 1238


RESULT 13
AR085078
LOCUS       AR085078               3221 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 1 from patent US 5981274.
ACCESSION   AR085078
VERSION     AR085078.1  GI:10011849
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3221)
  AUTHORS   Tyrrell,D.Lorne.J., Chaisomchit,S. and Chang,L.-J.
  TITLE     Recombinant hepatitis virus vectors
  JOURNAL   Patent: US 5981274-A 1 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .3221
                     /organism="unknown"
BASE COUNT      740 a    869 c    708 g    904 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 6;  Length 3221;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1177

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1237

Qy     123 GCG 125
          |||
Db    1238 GCG 1240


RESULT 14
E00010
LOCUS       E00010                 3221 bp    DNA     linear   PAT 29-SEP-1997
DEFINITION  DNA of hepatitis B virus (HBV).
ACCESSION   E00010
VERSION     E00010.1  GI:2168319
KEYWORDS    JP 1981063995-A/1.
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3221)
  AUTHORS   Uiriamu,J.R. and Hawaado,M.G.
  TITLE     NONTRANSIT VIRUS
  JOURNAL   Patent: JP 1981063995-A 1 30-MAY-1981;
            UNIV CALIFORNIA
COMMENT     OS   hepatitis B virus
            PN   JP 1981063995-A/1
            PD   30-MAY-1981
            PF   24-MAY-1980 JP 1980069516
            PR   24-MAY-1979 US 79     41909, 26-DEC-1979 US 79    107267 PI
            UIRIAMU JIEE RATSUTAA, HAWAADO MAIKERU GUTSUDOMAN PC
            C07H21/00,A61K39/29,C07C103/52,C07G7/00,C12N1/20,C12N15/00, PC
            C12P19/34//
            PC   C12R1/19,C12R1/91;
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No.
FEATURES             Location/Qualifiers
     source          1. .3221
                     /organism="Hepatitis B virus"
                     /db_xref="taxon:10407"
BASE COUNT      740 a    872 c    705 g    904 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 6;  Length 3221;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1119 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1178

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1179 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1238

Qy     123 GCG 125
          |||
Db    1239 GCG 1241


RESULT 15
AB014370
LOCUS       AB014370               3221 bp    DNA     circular VRL 06-FEB-1999
DEFINITION  Hepatitis B virus genomic DNA, complete sequence, isolate 11D11HCC.
ACCESSION   AB014370
VERSION     AB014370.1  GI:3551314
KEYWORDS    .
SOURCE      Hepatitis B virus (isolate:11D11HCC) DNA.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (sites)
  AUTHORS   Takahashi,K., Akahane,Y., Hino,K., Ohta,Y. and Mishiro,S.
  TITLE     Hepatitis B virus genomic sequence in the circulation of
            hepatocellular carcinoma patients: comparative analysis of 40
            full-length isolates
```

```
  REMARK     GenBank staff at the National Library of Medicine created this
             entry [NCBI gibbsq 138131] from the original journal article.
             This sequence comes from Fig. 5.
COMMENT      Strong homology to human hepatitis b virus.
FEATURES             Location/Qualifiers
     source          1. .3218
                     /organism="Duck hepatitis B virus"
                     /specific_host="duck"
                     /db_xref="taxon:12639"
     CDS             157. .837
                     /gene="Pre-S/S"
                     /note="This sequence comes from Fig. 5"
                     /codon_start=1
                     /product="preS/S"
                     /protein_id="AAL62458.1"
                     /db_xref="GI:18158616"
                     /translation="MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     SPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGTTTTSTGPCKTCTTPAQGNSMFPSCCCTKPSDGNCTCIPIPSSWAFAK
                     YLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVSPFIPLLPI
                     FFCLWVYI"
     gene            157. .834
                     /gene="Pre-S/S"
     gene            1376. .1840
                     /gene="X"
     CDS             1376. .1840
                     /gene="X"
                     /note="This sequence comes from Fig. 5"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAL62459.1"
                     /db_xref="GI:18158617"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLGLPA
                     MSTTDLEAYFKDCVFKDWEELGEEIRLKVFVLGGCRHKLVCAPAPCNFFTSA"
     gene            complement(1622. .2309)
                     /gene="P"
                     /pseudo
     CDS             complement(1622. .2309)
                     /gene="P"
                     /note="deletions at positions 2372 and 3028 resulting in
                     stop codons and frame shift"
                     /pseudo
                     /codon_start=1
     gene            1903. .2475
                     /gene="C"
                     /pseudo
     CDS             1903. .2475
                     /gene="C"
                     /note="stop codon within sequence"
                     /pseudo
                     /codon_start=1
BASE COUNT      739 a    857 c    714 g    908 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3218;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1177

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1237

Qy     123 GCG 125
           |||
Db    1238 GCG 1240


RESULT 12
AF462041
LOCUS       AF462041               3220 bp    DNA     circular VRL 01-FEB-2002
DEFINITION  Hepatitis B virus clone pAM6, complete genome.
ACCESSION   AF462041
VERSION     AF462041.1  GI:18389985
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3220)
  AUTHORS   Moriarty,A.M., Hoyer,B.H., Shih,J.W., Gerin,J.L. and Hamer,D.H.
  TITLE     Expression of the hepatitis B virus surface antigen gene in cell
            culture by using a simian virus 40 vector
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 78 (4), 2606-2610 (1981)
  MEDLINE   81223930
   PUBMED   6264484
REFERENCE   2  (bases 1 to 3220)
  AUTHORS   Jang,W.H., Yang,Y.I. and Kim,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2001) The Paik-Inje Memorial Institute for
            Biomedical Science, Inje University, 633-165 Gaegum-dong,
            Busanjin-gu, Busan 614-735, Korea
FEATURES             Location/Qualifiers
     source          1. .3220
                     /organism="Hepatitis B virus"
                     /db_xref="ATCC:45020"
                     /db_xref="taxon:10407"
                     /clone="pAM6; ATCC 45020"
                     /note="subtype: adw"
     gene            join(2306. .3220,1. .1623)
                     /gene="P"
     CDS             join(2306. .3220,1. .1623)
                     /gene="P"
                     /codon_start=1
                     /product="polymerase"
                     /protein_id="AAL68823.1"
                     /db_xref="GI:18389989"
                     /translation="MPLSYQHFRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELHHGRLVIKTSQRHGDESFCSQSSGILSRSSVGPCIRS
                     QLKQSRLGLQPHQGPLASSQPGRSGSIRARVHPSTRRCFGVEPSGSGHIDHSVNNSSS
                     CLHQSAVRKAAYSHLSTSKRQSSSGHAVELHCLPPSSAGSQSQGSVFSCWWLQFRNSK
                     PCSEYCLSHLVNLREDWGPCDEHGEHHIRIPRTPARVTGGVFLVDKNPHNTAESRLVV
                     DFSQFSRGITRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLL
                     IGSSGLSRYVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSH
                     PIVLGFRKIPMGVGLSPFLLAQFTSAICSVVRRAFPHCLAFSYMDDVVLGAKSVQHRE
                     SLYTAVTNFLLSLGIHLNPNKTKRWGYSLNFMGYVIGSWGTLPQDHIVQKIKHCFRKL
                     PVNRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQ
                     YMNLYPVARQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLPIHTAELLAACFARS
                     RSGAKLIGTDNSVVLSRKYTSFPWLLGCTANWILRGTSFVYVPSALNPADDPSRGRLG
                     LSRPLLRLPFQPTTGRTSLYAVSPSVPSHLPVRVHFASPLHVAWRPP"
     gene            join(2853. .3220,1. .835)
                     /gene="S"
     CDS             join(2853. .3220,1. .835)
                     /gene="S"
                     /note="pre-S1, pre-S2, S"
                     /codon_start=1
                     /product="surface antigen"
                     /protein_id="AAL68820.1"
                     /db_xref="GI:18389986"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNPDWD
                     FNPIKDHWPAANQVGVGAFGPGFTPPHGGVLGWSPQAQGILTTVSTIPPPASTNRQSG
                     RQPTPISPPLRDSHPQAMQWNSTAFHQALQDPRVRGLYFPAGGSSSGTVNPAPNIASH
                     ISSISARTGDPVTNMENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     SPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGSTTTSTGPCKTCTTPAQGNSMFPSCCCTKPTDGNCTCIPIPSSWAFAK
                     YLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVSPFIPLLPI
                     FFCLWVYI"
     gene            1374. .1838
                     /gene="X"
     CDS             1374. .1838
                     /gene="X"
```

```
Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2392 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2451

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2452 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2511

Qy     123 GCG 125
           |||
Db    2512 GCG 2514


RESULT 10
AF143306
LOCUS       AF143306               3137 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone RM517, complete genome.
ACCESSION   AF143306
VERSION     AF143306.1  GI:5019974
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3137)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 3137)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .3137
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="RM517"
                     /note="amplified by PCR; contains a premature termination
                     codon in C gene, a deletion in pre-S1/2 region removing
                     the pre-S2 start codon, a premature termination codon in S
                     gene, and a duplication plus insertion in X gene/core
                     promoter"
     gene            1. .171
                     /gene="pre-C/C"
     CDS             1. .171
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37953.1"
                     /db_xref="GI:5019975"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEL
                     LSFLPSDFFPSV"
     primer_bind     1. .28
                     /gene="pre-C/C"
     gene            494. .2935
                     /gene="P"
     CDS             494. .2935
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37954.1"
                     /db_xref="GI:5019976"
                     /translation="MPLSYQHLRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELQHGRLVIKTSQRHGDESFCSQPSGILSRSSVGPCIRS
                     QLKQSRLGLQPHQGPLASSQPGRSGSIRARAHPSTRRYFGVEPSGSGHIDHSVNEFHC
                     LPPSSAGSQSQGSVFSCWWLQFRNSKPCSEYCLSHLVNLREDWGPCGEHGEHHIRIPR
                     TPARVTGGVFLVDKNPHNTAESRLVVDFSQFSRGSTRVSWPKFAVPNLQSLTNLLSSN
                     LSWLSLDVSAAFYHIPLHPAAMPHLLIGSSGLSRYVARLSSNSRINNNQYGNMQNLHD
                     SCSRQLYVSLMLLYKTYGWKLHLYSHPIVLGFRKIPMGVGLSPFLLAQFTSAICSVVR
                     RAFPHCLAFSYMDDVVLGAKSVQHREALYTAVTNFLLSLGIHLNPNKTKRWGYSLNFM
                     GYIIGSWGTLPQDHIVQKIKHCFRELPVNRPIDWKVCQRIVGLLGFAAPFTQCGYPAL
                     MPLCACIQAKQAFTFSPTYKAFLSKQYMNLYPVARQRPGLCQVFADATPTGWGLAIGH
                     QRMRGTFVAPLPIHTAELLAACFARSRSGAKLIGTDNSVVLSRKYTSFPWLLGCTANW
                     ILRGTSFVYVPSALNPADDPSRGRLGLYRPLLRLPFQPTTGRTSLYAVSPSVPSHLPV
                     RVHFASPLHVAWRPP"
     gene            1041. .1751
                     /gene="pre-S/S"
     CDS             1041. .1751
                     /gene="pre-S/S"
                     /codon_start=1
                     /product="pre-S/S protein"
                     /protein_id="AAD37955.1"
                     /db_xref="GI:5019977"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNPDWD
                     FNPIKDHWPAANQVGVGAFGPGLTPPHGGILGWSPQAQGILTTVSTNSTAFHQALQDP
                     RVRGLYFPAGGSSSGTVNPAPNIASHISSISARTGDPVANMENITSGFLGPQLVLQAG
                     FFLLTRILTIPQSLDSWWTSLNFLGGAPVCLGQNSQSPTSNHSPTSCPPICPGYRWMC
                     LRRFIIFLFILLLCLIFL"
     gene            2686. .2979
                     /gene="X"
     CDS             2686. .2979
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37956.1"
                     /db_xref="GI:5019978"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPNPAQGLT"
     primer_bind     3130. .3137
BASE COUNT      723 a    835 c    693 g    886 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3137;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2428 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2487

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2488 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2547

Qy     123 GCG 125
           |||
Db    2548 GCG 2550


RESULT 11
S65868
LOCUS       S65868                 3218 bp    DNA     linear   VRL 16-JAN-2002
DEFINITION  Pre-S/S. . .C [duck hepatitis B virus pDKHBV, host=ducks, Genomic
            Mutant, 4 genes, 3218 nt].
ACCESSION   S65868
VERSION     S65868.1  GI:436272
KEYWORDS    .
SOURCE      Duck hepatitis B virus.
  ORGANISM  Duck hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
REFERENCE   1  (bases 1 to 3218)
  AUTHORS   Dai,W.L., Chen,Y., Li,L., Jiang,H.Q. and Gu,J.R.
  TITLE     Nucleotide sequence of a cloned human HBV mutant (pDKHBV) in duck
            hepatoma of Qidong County
  JOURNAL   Sci. China, Ser. B, Chem. Life Sci. Earth Sci. 36 (3), 329-338
            (1993)
  MEDLINE   94000357
   PUBMED   8397804
```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 21, 2003, 02:36:07 ; Search time 1023 Seconds
(without alignments)
3669.855 Million cell updates/sec

Title: US-09-689-430-1_COPY_150_278
Perfect score: 129
Sequence: 1 ctctttctaagtaaacagta.....gccatcagcgcatgcggatc 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	123	95.3	587	6	AR003584	AR003584 Sequence
2	123	95.3	587	6	AR062870	AR062870 Sequence
3	123	95.3	909	6	AR165345	AR165345 Sequence
4	123	95.3	1371	14	HPBVSAG2	M54898 Hepatitis B
5	123	95.3	2952	14	AF143307	AF143307 Hepatitis
6	123	95.3	3033	14	AF143299	AF143299 Hepatitis
7	123	95.3	3046	14	AF143301	AF143301 Hepatitis
8	123	95.3	3046	14	AF143308	AF143308 Hepatitis
9	123	95.3	3100	14	AF143305	AF143305 Hepatitis
10	123	95.3	3137	14	AF143306	AF143306 Hepatitis
11	123	95.3	3218	14	S65868	S65868 Pre-S/S...C
12	123	95.3	3220	14	AF462041	AF462041 Hepatitis
13	123	95.3	3221	6	AR085078	AR085078 Sequence
14	123	95.3	3221	6	E00010	E00010 DNA of hepa
15	123	95.3	3221	14	AB014370	AB014370 Hepatitis
16	123	95.3	3221	14	AB064314	AB064314 Hepatitis
17	123	95.3	3221	14	HBV012207	AJ012207 Hepatitis
18	123	95.3	3221	14	HBVADW2	X02763 Hepatitis b
19	123	95.3	3221	14	HBVXCPS	X70185 Hepatitis B
20	123	95.3	3221	14	HEB309369	AJ309369 Hepatitis
21	123	95.3	3221	14	HEB309370	AJ309370 Hepatitis
22	123	95.3	3221	14	HEB309371	AJ309371 Hepatitis
23	123	95.3	3221	14	HUMPRECX	L13994 Hepatitis B
24	123	95.3	4084	12	AF305422	AF305422 Synthetic
c 25	123	95.3	4525	6	AR062871	AR062871 Sequence
26	123	95.3	4627	6	I03789	I03789 Sequence 4
27	123	95.3	6371	6	AR085082	AR085082 Sequence
28	123	95.3	6371	6	AR085089	AR085089 Sequence
29	123	95.3	6375	6	AR085091	AR085091 Sequence
30	123	95.3	9325	6	AR085079	AR085079 Sequence
31	123	95.3	9354	6	AR003585	AR003585 Sequence
32	123	95.3	9859	6	AR085083	AR085083 Sequence
33	121.4	94.1	1355	6	E00070	E00070 DNA coding
34	121.4	94.1	1355	6	E00121	E00121 DNA coding
35	121.4	94.1	3058	14	AF143300	AF143300 Hepatitis
36	121.4	94.1	3125	14	AF143302	AF143302 Hepatitis
37	121.4	94.1	3126	14	AF143304	AF143304 Hepatitis
38	121.4	94.1	3161	14	AF143298	AF143298 Hepatitis
39	121.4	94.1	3183	6	E00120	E00120 DNA of adw
40	121.4	94.1	3200	6	E01301	E01301 DNA of hepa
41	121.4	94.1	3200	6	E10905	E10905 Genomic DNA
42	121.4	94.1	3200	14	AF297620	AF297620 Hepatitis
43	121.4	94.1	3200	14	AF297624	AF297624 Hepatitis
44	121.4	94.1	3200	14	HBVADW	V00866 Hepatitis B
45	121.4	94.1	3221	14	AF090838	AF090838 Hepatitis

ALIGNMENTS

RESULT 1
AR003584
LOCUS AR003584 587 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5744326.
ACCESSION AR003584
VERSION AR003584.1 GI:3964843
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 587)
AUTHORS Ill,C.R. and Bidlingmaier,S.
TITLE Use of viral CIS-acting post-transcriptional regulatory sequences to increase expression of intronless genes containing near-consensus splice sites

```
  JOURNAL   Patent: US 5744326-A 1 28-APR-1998;
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="unknown"
BASE COUNT       97 a    199 c    145 g    146 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 6;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 61

Qy     63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 121

Qy    123 GCG 125
          |||
Db    122 GCG 124


RESULT 2
AR062870
LOCUS       AR062870                 587 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5843770.
ACCESSION   AR062870
VERSION     AR062870.1  GI:5990561
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Ill,C.R. and Gonzales,J.E.N.
  TITLE     Antisense constructs directed against viral post-transcriptional
            regulatory sequences
  JOURNAL   Patent: US 5843770-A 1 01-DEC-1998;
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="unknown"
BASE COUNT       97 a    199 c    145 g    146 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 6;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 61

Qy     63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 121

Qy    123 GCG 125
          |||
Db    122 GCG 124


RESULT 3
AR165345
LOCUS       AR165345                 909 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6274788.
ACCESSION   AR165345
VERSION     AR165345.1  GI:16238916
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Kumar,V., Singh,M.; Totey,S. and Anand,R.
  TITLE     Bicistronic DNA construct comprising X-myc transgene for use in
            production of transgenic animal model systems for human
            hepatocellular carcinoma and transgenic animal model systems so
            produced
  JOURNAL   Patent: US 6274788-A 1 14-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .909
                     /organism="unknown"
BASE COUNT      210 a    236 c    211 g    252 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 6;  Length 909;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 345

Qy     63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 405

Qy    123 GCG 125
          |||
Db    406 GCG 408


RESULT 4
HPBVSAG2
LOCUS       HPBVSAG2                1371 bp    DNA     circular VRL 06-MAR-1995
DEFINITION  Hepatitis B virus surface protein gene,complete cds.
ACCESSION   M54898 M38545
VERSION     M54898.1  GI:329731
KEYWORDS    Dane particle protein; S protein; hepatitis B surface antigen;
            virion protein.
SOURCE      Human (hepatitis B carrier from Taiwan) virion (Dane particle) DNA,
            clone pTWS1.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 1371)
  AUTHORS   Ting,L.-P. and Chang,H.-K.
  TITLE     Characterization of nucleotide seqence of hepatitis B surface gene
            and enhancer element of HBV
  JOURNAL   Chinese J. Microbiol. Immunol. 20, 224-240 (1987)
FEATURES             Location/Qualifiers
     source          1. .1371
                     /organism="Hepatitis B virus"
                     /db_xref="taxon:10407"
                     /clone="pTWS1"
     CDS             124. .804
                     /codon_start=1
                     /product="surface protein"
                     /protein_id="AAA62816.1"
                     /db_xref="GI:329732"
                     /translation="MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     SPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGTTTTSTGPCKTCTTPAQGNSMFPSCCCTKPSDGNCTCIPIPSSWAFAK
                     YLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVSPFIPLLPI
                     FFCLWVYI"
BASE COUNT      301 a    359 c    301 g    410 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 1371;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1085 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1144

Qy     63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1145 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1204

Qy     123 GCG 125
           |||
Db    1205 GCG 1207


RESULT 5
AF143307
LOCUS       AF143307               2952 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone RM518, complete genome.
ACCESSION   AF143307
VERSION     AF143307.1  GI:5019979
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 2952)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 2952)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .2952
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="RM518"
                     /note="amplified by PCR; contains a deletion in C gene, a
                     deletion in pre-S2 region, a premature termination codon
                     in S gene, and a deletion in X gene/core promoter"
     gene            1. .408
                     /gene="pre-C/C"
     CDS             1. .408
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37957.1"
                     /db_xref="GI:5019980"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEL
                     LSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRDAPILSRLPETTVVRRR
                     DRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC"
     primer_bind     1. .28
                     /gene="pre-C/C"
     gene            257. .2770
                     /gene="P"
     CDS             257. .2770
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37958.1"
                     /db_xref="GI:5019981"
                     /translation="MPLSYRDFRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIIDRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELQHGRLVIKTSQRHGDESFCSQPAGILSRSSVGPCIRS
                     QLKQSRLGLQPHQGPLASSQPGRSGSIRARAHPSTRRYFGVEPSGSGHIDHSVNNSSS
                     CLHQSAVRKAAYSHLSTSKRQSSSGHASSAGSQSQGSAFSCWWLQFRNSKPCSEYCLS
                     HLVNLREDWGPCAEHGEHHIRIPRTPARVTGGVFLVDKNPHNTAESRLVVDFSQFSRG
                     ITRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLLVGSSGLSR
                     YVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSHPIVLGFRK
                     IPMGVGLSPFLLAQFTSAICSVVRRAFPHCLAFSYMDDVVLGAKSVQHRESLYTAVTN
                     FLLSLGIHLNPNKTKRWGYSLNFMGYIIGSWGTLPQDHIVQKIKHCFRKLPVNRPIDW
                     KVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQYMNLYPVA
                     RQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLPIHTAELLAACFARSRSGAKLIG
                     TDNSVVLSRKYTSFPWLLGCTANWILRGTSFVYVPSALNPADDPSRGRLGLYRPLLRL
                     PFQPATGRTSLYAVSPSVPSHLPVRVHFASPLHVAWRPP"
     gene            804. .1586
                     /gene="pre-S/S"
     CDS             804. .1586
                     /gene="pre-S/S"
                     /codon_start=1
                     /product="pre-S/S protein"
                     /protein_id="AAD37959.1"
                     /db_xref="GI:5019982"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNLDWD
                     FNPIKDHWPAANQVGVGAFGPGLTPPHGGILGWSPQAQGILTTVSTIPPPASTNRQSG
                     RQPTPISPPLRDSHPQAMQALQDPRVRGLHFPAGGSSSGTVNPAPNIASHISSISART
                     GDPVPNMENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGSPVCLGQN
                     SQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFL"
     gene            2521. .2925
                     /gene="X"
     CDS             2521. .2925
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37960.1"
                     /db_xref="GI:5019983"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPAGHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLGLPA
                     MSTTDLEAYFKDCVFKDWEELGEEIRLNIRRL"
     primer_bind     2945. .2952
BASE COUNT      675 a    808 c    651 g    818 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 2952;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2263 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2322

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2323 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2382

Qy     123 GCG 125
           |||
Db    2383 GCG 2385


RESULT 6
AF143299
LOCUS       AF143299               3033 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone 7744-G9, complete genome.
ACCESSION   AF143299
VERSION     AF143299.1  GI:5019937
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3033)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 3033)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .3033
```

```
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="7744-G9"
                     /note="amplified by PCR; contains a deletion in C gene, a
                     deletion in pre-S1/2 region removing the pre-S2 start
                     codon, a premature termination codon in pre-S1 region, and
                     a deletion in X gene/core promoter"
     gene            1. .543
                     /gene="pre-C/C"
     CDS             1. .543
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37923.1"
                     /db_xref="GI:5019938"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEL
                     LSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVG
                     NNLGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTPPETTVVRRRDRGRSPRRRTPSP
                     RRRRSQSPRRRRSQSRESQC"
     primer_bind     1. .28
                     /gene="pre-C/C"
     gene            392. .2851
                     /gene="P"
     CDS             392. .2851
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37924.1"
                     /db_xref="GI:5019939"
                     /translation="MPLSYQHLRRLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELQHGRLVIKTSQRHGDESFCSQPSGILSRSSVGPCIRS
                     QLKQSRLGLQPHQGPLASSQPGRSGSIRARAHPSTRRYLGVEPSGSGHMDHSVNNSSS
                     CLHQSAVPPSSAGSQSQGSVFSCWWLQFRNSKPCSEYCLSHLVNLREDWGPCDEHGEH
                     HIRIPRTPARVTGGVFLVDKNPHNTTESRLVVDFSQFSRGITRVSWPKFAVPNLQSLT
                     NLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLLIGSSGLSRYVARLSSNSRINNNQYGT
                     MQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSHPIVLGFRKIPMGVGLSPFLLAQFTSA
                     ICSVVRRAFPHCLAFSYMDDVVLGAKSVQHRESLYTAVTNFLLSLGIHLNPNKTKRWG
                     YSLNFMGYIIGSWGTLPQDHIVQKIKHCFRKLPVNRPIDWKVCQRIVGLLGFAAPFTQ
                     CGYPALMPLYACIQAKQAFTFSPTYKAFLSKQYMNLYPVARQRPGLCQVFADATPTGW
                     GLAIGHQRMRGTFVAPLPIHTAELLAACFARSRSGAKLIGTDNSVVLSRKYTSFPWLL
                     GCTANWILRGTSFVYVPSALNPADDPSRGRLGLYRPLLRLPFQPTTGRTSLYAVSPSV
                     PSHLPVRVHFASPLHVAWRPP"
     gene            939. .1163
                     /gene="pre-S1"
     CDS             939. .1163
                     /gene="pre-S1"
                     /codon_start=1
                     /product="pre-S1 protein"
                     /protein_id="AAD37925.1"
                     /db_xref="GI:5019940"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNPDWD
                     FNPIKDHWPAANQVGVGAFGPGLTPPHGGI"
     gene            1383. .2063
                     /gene="S"
     CDS             1383. .2063
                     /gene="S"
                     /codon_start=1
                     /product="S protein"
                     /protein_id="AAD37926.1"
                     /db_xref="GI:5019941"
                     /translation="MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     SPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGSTTTSTGPCKTCMTPAQGNSMFPSCCCTKPTDGNCTCIPIPSSWAFAK
                     YLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVSPFIPLLPI
                     FFCLWVYI"
     gene            2602. .3006
                     /gene="X"
     CDS             2602. .3006
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37927.1"
                     /db_xref="GI:5019942"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLGLPA
                     MSTTDLEAYFKDCVFKDWEELGEEIRLNIRRL"
     primer_bind     3026. .3033
BASE COUNT      695 a    820 c    674 g    844 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3033;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2344 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2403

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2404 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2463

Qy     123 GCG 125
          |||
Db    2464 GCG 2466


RESULT 7
AF143301
LOCUS       AF143301               3046 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone BW1903, complete genome.
ACCESSION   AF143301
VERSION     AF143301.1  GI:5019947
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3046)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 3046)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .3046
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="BW1903"
                     /note="amplified by PCR; contains a deletion in pre-S1
                     region, a premature termination codon in S gene, and a
                     duplication plus a deletion in the X gene/core promoter"
     gene            1. .645
                     /gene="pre-C/C"
     CDS             1. .645
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37931.1"
                     /db_xref="GI:5019948"
                     /translation="MQLFHLCLIISCSCPTVQASKLCLGWLWGMDIDTYKEFGATVEL
                     LSFLPSDFFPSIRDLVDTASALHREALESPEHCSPHHTALRQAIVCWGELMTLATWVG
                     NNLEDPASRDLVVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPA
                     YRPPNAPILSTLPETTVVRRRDRGRSPRRRTPSPRRRRSQSPRRRRSQSRESQC"
     primer_bind     1. .28
                     /gene="pre-C/C"
```

```
     gene            494. .2836
                     /gene="P"
     CDS             494. .2836
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37932.1"
                     /db_xref="GI:5019949"
                     /translation="MPLSYQHFRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRGTTRSASFCGSPYSWEQELQHGRLVTKTSQRHGDESFCSQPSGILSRSSVGHQSAV
                     RKAAYSHFSTSKRQSSSGHAVEFHCLPPSSAGSQSQGSVFSCWWLQFRNSKPCSEYCL
                     SHLVNLREDWGPCGEHGEHHIRIPRTPARVTGGVFLVDKNPHNTAESRLVVDFSQFSR
                     GSTRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLLIGSSGLS
                     RYVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSHPIVLGFR
                     KIPMGVGLSPFLLAQFTSAICSVIRRAFPHCLAFSYMDDVVLGAKSVQHRESLYTAVT
                     NFLLSLGIHLNPNKTKRWGYSLNFMGYIIGSWGTLPQDHIVQKIKHCFRKLPVNRPID
                     WKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQYMNLYPV
                     ARQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLPIHTAELLAACFARSRSGAKLI
                     GTDNSVVLSRKYTSFPWLLGCAANWILRGTSFVYVPSALNPADDPSRGRLGLYRPLLR
                     LPFQPTTGRTSLYAVSPSVPSHLPVRVHFASPLHVAWRPP"
     gene            1041. .1913
                     /gene="pre-S/S"
     CDS             1041. .1913
                     /gene="pre-S/S"
                     /codon_start=1
                     /product="pre-S/S protein"
                     /protein_id="AAD37933.1"
                     /db_xref="GI:5019950"
                     /translation="MGGWSPKPRKGMGTNLSVPNPLGFFPDHQLATNRQSGRQPTPIS
                     PPLRDSHPQAMQWNSTAFHQALQDPKVRGLYFPAGGSSSGTVNPAPNIASHISSISAR
                     TGDPVANMENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGAPVCLGQ
                     NSQSPTSNHSPTSCPPICPGYRWMCLQRFIIFLFILLLCLIFLLVLLDYQGMLPVCPL
                     IPGSTTTSTGPCKTCTTPAQGNSMFPSCCCTKPTDGNCTCIPIPSSWAFAKYLWEWAS
                     VRFSWLSLLVPFVQ"
     gene            2587. .2898
                     /gene="X"
     CDS             2587. .2898
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37934.1"
                     /db_xref="GI:5019951"
                     /translation="MAARLCCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKRTLAQCL
                     T"
     primer_bind     3039. .3046
BASE COUNT      708 a    803 c    663 g    872 t
ORIGIN

  Query Match             95.3%;  Score 123;  DB 14;  Length 3046;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2329 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2388

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2389 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2448

Qy     123 GCG 125
          |||
Db    2449 GCG 2451


RESULT 8
AF143308
LOCUS       AF143308               3046 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone WB1254, complete genome.
ACCESSION   AF143308
VERSION     AF143308.1  GI:5019984
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3046)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 3046)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .3046
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="WB1254"
                     /note="amplified by PCR; contains a premature termination
                     codon in C gene, a deletion in pre-S1 region, a premature
                     termination codon in S gene, and a duplication plus a
                     deletion in the X gene/core promoter"
     gene            1. .639
                     /gene="pre-C/C"
     CDS             1. .639
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37961.1"
                     /db_xref="GI:5019985"
                     /translation="MQLFHLCLIISCSCPTVQASKLCLGWLWGMDIDTYKEFGATVEL
                     LSFLPSDFFPSIRDLVDTASALHREALESPEHCSPHHTALRQAIVCWGELMTLATWVG
                     NNLEDPASRDLVVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPA
                     YRPPNAPILSTLPETAVVRRRDRGRSPRRRTPSPRRRRSQSPRRRRSQSRES"
     primer_bind     1., .28
                     /gene="pre-C/C"
     gene            494. .2836
                     /gene="P"
     CDS             494. .2836
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37962.1"
                     /db_xref="GI:5019986"
                     /translation="MPLSYQHFRKLLLLDDGTEAGPLEEELPRLADADLNRRVAEDLN
                     LGNLNVSIPWTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLT
                     VNEKRRLKLIMPARFHPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILY
                     KRETTRSASFCGSPYSWEQELQHGRLVTKTSQRHGDESFCSQPSGILSRSSVGHQSAV
                     RKAAYSHFSTSKRQSSSGHAVEFHCLPPSSAGSQSQGSVSSCWWLQFRNSKPCSEYCL
                     SHLVNLREDWGPCGEHGEHHIRIPRTPARVTGGVFLVDKNPHNTAESRLVVDFSQFSR
                     GSTRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAFYHIPLHPAAMPHLLIGSSGLS
                     RYVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLMLLYKTYGWKLHLYSHPIVLGFR
                     KIPMGVGLSPFLLAQFTSAICSVIRRAFPHCLAFSYMDDVVLGAKSVQHRESLYTAVT
                     NFLLSLGIHLNPNKTKRWGYSLNFMGYIIGSWGTLPQDHIVQKIKHCFRKLPVNRPID
                     WKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQAFTFSPTYKAFLSKQYMNLYPV
                     ARQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLPIHTAELLAACFARSRSGAKLI
                     GTDNSVVLSRKYTSFPWLLGCTANWILRGTSFVYVPSALNPADDPSRGRLGLYRPLLR
                     LPFQPTTGRTSLYAVSPSVPSHLPVRVHFASPLHVAWRPP"
     gene            1041. .1652
                     /gene="pre-S/S"
     CDS             1041. .1652
                     /gene="pre-S/S"
                     /codon_start=1
                     /product="pre-S/S protein"
                     /protein_id="AAD37963.1"
                     /db_xref="GI:5019987"
                     /translation="MGGWSPKPRKGMGTNLSVPNPLGFFPDHQLATNRQSGRQPTPIS
```

```
                     PPLRDSHPQAMQWNSTAFHQALQDPRVRGLYLPAGGSSSGTVNPAPNIASHISSISAR
                     TGDPVANMENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGAPVCLGQ
                     NSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFL"
     gene            2587. .2898
                     /gene="X"
     CDS             2587. .2898
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37964.1"
                     /db_xref="GI:5019988"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKVLHKGTLAQCL
                     T"
     primer_bind     3039. .3046
BASE COUNT      704 a    806 c    666 g    870 t
ORIGIN

  Query Match           95.3%;  Score 123;  DB 14;  Length 3046;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2329 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2388

Qy     63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2389 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 2448

Qy    123 GCG 125
          |||
Db   2449 GCG 2451


RESULT 9
AF143305
LOCUS       AF143305               3100 bp    DNA     circular VRL 19-OCT-1999
DEFINITION  Hepatitis B virus clone RM501, complete genome.
ACCESSION   AF143305
VERSION     AF143305.1  GI:5019968
KEYWORDS    .
SOURCE      Hepatitis B virus.
  ORGANISM  Hepatitis B virus
            Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
REFERENCE   1  (bases 1 to 3100)
  AUTHORS   Preikschat,P., Meisel,H., Will,H. and Gunther,S.
  TITLE     Hepatitis B virus genomes from long-term immunosuppressed virus
            carriers are modified by specific mutations in several regions
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2685-2691 (1999)
  MEDLINE   20037832
   PUBMED   10573161
REFERENCE   2  (bases 1 to 3100)
  AUTHORS   Preikschat,P., Meisel,H., Iwanska,A., Will,H. and Gunther,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Department of Virology,
            Bernhard-Nocht-Institute for Tropical Medicine,
            Bernhard-Nocht-Strasse 74, Hamburg D-20359, Germany
FEATURES             Location/Qualifiers
     source          1. .3100
                     /organism="Hepatitis B virus"
                     /virion
                     /db_xref="taxon:10407"
                     /clone="RM501"
                     /note="amplified by PCR; contains a deletion in C gene/P
                     gene, a deletion in pre-S1/2 region removing the pre-S2
                     start codon, a premature termination codon in pre-S1
                     region, and a duplication plus deletion in X gene/core
                     promoter"
     gene            1. .618
                     /gene="pre-C/C"
     CDS             1. .618
                     /gene="pre-C/C"
                     /codon_start=1
                     /product="pre-C/C protein"
                     /protein_id="AAD37948.1"
                     /db_xref="GI:5019969"
                     /translation="MQLFHLCLIISCTCPTVQASKLCLGWLWGMDIDPYKEFGATVEF
                     LSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVG
                     NNLEDPASRDLVVNYVKTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPA
                     YRPPNAPILSTPPETTVVRRRNRGRSPRRRTPSRRRRSQSRESQC"
     primer_bind     1. .28
                     /gene="pre-C/C"
     gene            494. .2899
                     /gene="P"
     CDS             494. .2899
                     /gene="P"
                     /codon_start=1
                     /product="P protein"
                     /protein_id="AAD37949.1"
                     /db_xref="GI:5019970"
                     /translation="MPLSYQHLRKLLLLDDGTEAGPLEEELPRVAEDLNLGNLNVSIP
                     WTHKVGNFTGLYSSTVPIFNPEWQTPSFPKIHLQEDIINRCQQFVGPLTVNEKRRLKL
                     IMPARFYPTHTKYLPLDKGIKPYYPDQVVNHYFQTRHYLHTLWKAGILYKRETTRSAS
                     FCGSPYSWEQELQHGRLVIKTSQRHGDESFCSQPSGILSRSSVGPCIRSQLKQSRLGL
                     QPHQGQLASSQPGRSGSIRARAHPSTRRYFGVKPSGSGHIDHSVNCLPPSSAGSQSQG
                     SVFSCWWLQFRNSKPCSEYCLSHLINLREDWGPCDEHGEHHIRIPRTPARVTGGVFLV
                     DKNPHNTAESRLVVDFSQFSRGTTRVSWPKFAVPNLQSLTNLLSSNLSWLSLDVSAAF
                     YHIPLHPAAMPHLLIGSSGLSRYVARLSSNSRINNNQYGTMQNLHDSCSRQLYVSLML
                     LYKTYGWKLHLYSHPIVLGFRKIPMGVGLSPFLLAQFTSAICSVVRRAFPHCLAFSYM
                     DDVVLGAKSVQHREALYTAVTNFLLSLGIHLNPNKTKRWGYSLNFMGYIIGSWGTLPQ
                     DHIVQKIKHCFRKLPVNRPIDWKVCQRIVGLLGFAAPFTQCGYPALMPLYACIQAKQA
                     FTFSPTYKAFLSKQYMNLYPVARQRPGLCQVFADATPTGWGLAIGHQRMRGTFVAPLP
                     IHTAELLAACFARSRSGAKLVGTDNSVVLSRKYTSFPWLLGCTANWILRGTSFVYVPS
                     ALNPADDPSRGRLGLYRPLLRLPFQPTTGRTSLYAVSPSVPSHLPVRVHFASPLHVAW
                     RPP"
     gene            1014. .1244
                     /gene="pre-S1"
     CDS             1014. .1244
                     /gene="pre-S1"
                     /codon_start=1
                     /product="pre-S1 protein"
                     /protein_id="AAD37950.1"
                     /db_xref="GI:5019971"
                     /translation="MGGWSSKPRKGMGTNLSVPNPLGFFPDHQLDPAFGANSNNPDWD
                     FNPIKDNWPAANQVGVGAFGPGLTPPHGGILG"
     gene            1431. .2111
                     /gene="S"
     CDS             1431. .2111
                     /gene="S"
                     /codon_start=1
                     /product="S protein"
                     /protein_id="AAD37951.1"
                     /db_xref="GI:5019972"
                     /translation="MENITSGFLGPLLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGG
                     PPVCLGQNSQSPTSNHSPTSCPPICPGYRWMCLRRFIIFLFILLLCLIFLLVLLDYQG
                     MLPVCPLIPGSTTTSTGPCKTCTTPAQGNSMFPSCCCTKPTDGNCTCIPIPSSWAFAK
                     YLWEWASVRFSWLSLLVPFVQWFVGLSPTVWLSAIWMMWYWGPSLYSIVRPFIPLLPI
                     FFCLWVYI"
     gene            2650. .2952
                     /gene="X"
     CDS             2650. .2952
                     /gene="X"
                     /codon_start=1
                     /product="X protein"
                     /protein_id="AAD37952.1"
                     /db_xref="GI:5019973"
                     /translation="MAARLYCQLDPSRDVLCLRPVGAESRGRPLSGPLGTLSSPSPSA
                     VPADHGAHLSLRGLPVCAFSSAGPCALRFTSARCMETTVNAHQILPKALHKRTLLT"
     primer_bind     3093. .3100
BASE COUNT      723 a    821 c    682 g    874 t
ORIGIN

  Query Match           95.3%;  Score 123;  DB 14;  Length 3100;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-29;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 21, 2003, 03:48:37 ; Search time 105 Seconds
(without alignments)
1622.282 Million cell updates/sec

Title: US-09-689-430-1_COPY_150_278
Perfect score: 129
Sequence: 1 ctctttctaagtaaacagta.....gccatcagcgcatgcggatc 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	129	100.0	7914	12	US-10-095-718-3	Sequence 3, Appli
	2	129	100.0	7944	12	US-10-095-718-1	Sequence 1, Appli
	3	119.8	92.9	3221	9	US-09-848-616-133	Sequence 133, App
	4	110.2	85.4	306	9	US-09-875-453-25	Sequence 25, Appl
	5	110.2	85.4	3182	9	US-10-104-966-14	Sequence 14, Appl
	6	110.2	85.4	3182	10	US-09-929-955-14	Sequence 14, Appl
	7	110.2	85.4	5618	9	US-10-142-358-1	Sequence 1, Appli
	8	110.2	85.4	7991	10	US-09-837-297-5	Sequence 5, Appli
	9	110.2	85.4	8007	10	US-09-837-297-3	Sequence 3, Appli
	10	110.2	85.4	8717	10	US-09-837-297-4	Sequence 4, Appli
	11	105	81.4	3215	9	US-10-209-264-1	Sequence 1, Appli
	12	99	76.7	5130	9	US-09-897-511A-9	Sequence 9, Appli
	13	99	76.7	5130	10	US-09-897-006-9	Sequence 9, Appli
	14	52.4	40.6	67	9	US-09-466-035-55	Sequence 55, Appl
	15	52.4	40.6	67	10	US-09-912-679-55	Sequence 55, Appl
	16	51.2	39.7	592	9	US-10-202-457-1	Sequence 1, Appli
	17	51.2	39.7	5691	9	US-09-897-511A-11	Sequence 11, Appl
	18	51.2	39.7	5691	10	US-09-897-006-11	Sequence 11, Appl
	19	51.2	39.7	5711	9	US-09-897-511A-8	Sequence 8, Appli
	20	51.2	39.7	5711	10	US-09-897-006-8	Sequence 8, Appli
	21	51.2	39.7	5732	9	US-09-897-511A-6	Sequence 6, Appli
	22	51.2	39.7	5732	10	US-09-897-006-6	Sequence 6, Appli
	23	51.2	39.7	9183	9	US-09-897-511A-7	Sequence 7, Appli
	24	51.2	39.7	9183	10	US-09-897-006-7	Sequence 7, Appli
	25	51.2	39.7	9941	9	US-10-243-553-3	Sequence 3, Appli
	26	51.2	39.7	9941	9	US-10-243-816-2	Sequence 2, Appli
	27	34	26.4	1731	9	US-09-738-626-1078	Sequence 1078, Ap
c	28	34	26.4	3309400	9	US-09-738-626-1	Sequence 1, Appli
c	29	30.8	23.9	442	10	US-09-864-761-868	Sequence 868, App
c	30	30.8	23.9	462	10	US-09-864-761-796	Sequence 796, App
c	31	30.8	23.9	480	10	US-09-864-761-264	Sequence 264, App
	32	29	22.5	2469	10	US-09-853-386-90	Sequence 90, Appl
	33	29	22.5	3421	10	US-09-853-386-14	Sequence 14, Appl
	34	29	22.5	3427	10	US-09-853-386-4	Sequence 4, Appli
	35	29	22.5	9824	10	US-09-853-386-1	Sequence 1, Appli
	36	29	22.5	10883	10	US-09-853-386-13	Sequence 13, Appl
c	37	29	22.5	14136	10	US-09-964-824A-244	Sequence 244, App
c	38	29	22.5	53522	9	US-09-904-968A-1	Sequence 1, Appli
c	39	28.4	22.0	38155	9	US-10-114-170-79	Sequence 79, Appl
	40	27.8	21.6	5836	9	US-10-114-170-253	Sequence 253, App
c	41	27.2	21.1	356	10	US-09-867-701-419	Sequence 419, App
	42	27	20.9	2667	9	US-09-895-298-12	Sequence 12, Appl
	43	27	20.9	2714	9	US-10-227-884-21	Sequence 21, Appl
	44	27	20.9	2714	9	US-10-230-163-21	Sequence 21, Appl
	45	27	20.9	2714	9	US-10-218-631-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-095-718-3
; Sequence 3, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Burstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
; FILE REFERENCE: 35052/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7914
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: rAAV vector with canine B-domain deleted factor
; OTHER INFORMATION: VIII
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (435)...(4730)
US-10-095-718-3

Query Match 100.0%; Score 129; DB 12; Length 7914;
Best Local Similarity 100.0%; Pred. No. 8.5e-38;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 209
```

```
Qy      61 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 269

Qy     121 ATGCGGATC 129
           |||||||||
Db     270 ATGCGGATC 278


RESULT 2
US-10-095-718-1
; Sequence 1, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
;  APPLICANT: Walsh, Christopher
;  APPLICANT:  Chao, Hengjun
;  APPLICANT:  Burstein, Haim
;  APPLICANT:  Lynch, Carmel
;  APPLICANT:  Stepan, Tony
;  APPLICANT:  Munson, Keith
;  TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
;  TITLE OF INVENTION:  Methods of Using the Same
;  FILE REFERENCE: 35052/204375
;  CURRENT APPLICATION NUMBER: US/10/095,718
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: 09/689,430
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/158,780
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 7944
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pDLZ6 encoding Homo sapiens BDD FVIII
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (420)...(4835)
US-10-095-718-1

  Query Match             100.0%;  Score 129;  DB 12;  Length 7944;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-38;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 209

Qy      61 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 269

Qy     121 ATGCGGATC 129
           |||||||||
Db     270 ATGCGGATC 278


RESULT 3
US-09-848-616-133
; Sequence 133, Application US/09848616
; Publication No. US20030054010A1
; GENERAL INFORMATION:
;  APPLICANT: Sebbel, Peter
;  APPLICANT:  Dunant, Nicolas
;  APPLICANT:  Bachmann, Martin
;  APPLICANT:  Tissot, Alain
;  APPLICANT:  Lechner, Franziska
;  TITLE OF INVENTION: Molecular Antigen Array
;  FILE REFERENCE: 1700.0180002
;  CURRENT APPLICATION NUMBER: US/09/848,616
;  CURRENT FILING DATE:  2001-05-05
;  NUMBER OF SEQ ID NOS: 186
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 133
;   LENGTH: 3221
;   TYPE: DNA
;   ORGANISM: Hepatitis B virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1901)..(2458)
US-09-848-616-133

  Query Match             92.9%;  Score 119.8;  DB 9;  Length 3221;
  Best Local Similarity   98.4%;  Pred. No. 1.7e-34;
  Matches  121;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1116 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1175

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    1176 CCAAGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCTTGGCCATAGGCCATCAGCGCAT 1235

Qy     123 GCG 125
           | |
Db    1236 GAG 1238


RESULT 4
US-09-875-453-25
; Sequence 25, Application US/09875453
; Publication No. US20030027320A1
; GENERAL INFORMATION:
;  APPLICANT: Kim, Jungsuh P.
;  APPLICANT:  Starr, Douglas B.
;  APPLICANT:  Tam, Albert W.
;  APPLICANT:  Laurance, Megan E.
;  APPLICANT:  Michelotti, Emil F.
;  APPLICANT:  Velligan, Mark D.
;  APPLICANT:  Latour, Derek R.
;  APPLICANT:  Thomas, Rita L.
;  APPLICANT:  Kongpachith, Ana
;  APPLICANT:  Sheppard, Liana T.
;  APPLICANT:  Lim, Moon Young
;  APPLICANT:  Bruice, Thomas W.
;  TITLE OF INVENTION: PROMOTERS FOR REGULATED GENE EXPRESSION
;  FILE REFERENCE: 4600-0135.30
;  CURRENT APPLICATION NUMBER: US/09/875,453
;  CURRENT FILING DATE:  2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/209,549
;  PRIOR FILING DATE: 2000-06-06
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 306
;   TYPE: DNA
;   ORGANISM: Hepatitis B virus
US-09-875-453-25

  Query Match             85.4%;  Score 110.2;  DB 9;  Length 306;
  Best Local Similarity   93.5%;  Pred. No. 3.3e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  |||||||  ||| ||||||||||||||||||  |||||||||||  ||||||||
Db      38 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 97

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db      98 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 157

Qy     123 GCG 125
           |||
```

```
Db    158 GCG 160


RESULT 5
US-10-104-966-14
; Sequence 14, Application US/10104966
; Patent No. US20020155124A1
; GENERAL INFORMATION:
;  APPLICANT: Matti Sallberg
;  APPLICANT:  Catharina Hultgren
;  TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
;  TITLE OF INVENTION:  METHODS OF USE THEREOF
;  FILE REFERENCE: TRIPEP.23AUSC1
;  CURRENT APPLICATION NUMBER: US/10/104,966
;  CURRENT FILING DATE:  2002-03-22
;  PRIOR APPLICATION NUMBER: 09/705,547
;  PRIOR FILING DATE: 2000-11-03
;  PRIOR APPLICATION NUMBER: 60/229,175
;  PRIOR FILING DATE: 2000-08-29
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 3182
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Hepatitis B virus sequence
US-10-104-966-14

  Query Match             85.4%;  Score 110.2;  DB 9;  Length 3182;
  Best Local Similarity   93.5%;  Pred. No. 6.3e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    1118 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 1177

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    1178 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 1237

Qy     123 GCG 125
           |||
Db    1238 GCG 1240


RESULT 6
US-09-929-955-14
; Sequence 14, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
;  APPLICANT: Matti Sallberg
;  APPLICANT:  Catharina Hultgren
;  TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
;  TITLE OF INVENTION:  METHODS OF USE THEREOF
;  FILE REFERENCE: TRIPEP.23AUS2
;  CURRENT APPLICATION NUMBER: US/09/929,955
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/705,547
;  PRIOR FILING DATE: 2000-11-03
;  PRIOR APPLICATION NUMBER: 60/229,175
;  PRIOR FILING DATE: 2000-08-29
;  PRIOR APPLICATION NUMBER: 60/225,767
;  PRIOR FILING DATE: 2000-08-17
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 3182
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Hepatitis B virus sequence
US-09-929-955-14

  Query Match             85.4%;  Score 110.2;  DB 10;  Length 3182;
  Best Local Similarity   93.5%;  Pred. No. 6.3e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    1118 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 1177

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    1178 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 1237

Qy     123 GCG 125
           |||
Db    1238 GCG 1240


RESULT 7
US-10-142-358-1
; Sequence 1, Application US/10142358
; Publication No. US20030083291A1
;   GENERAL INFORMATION:
;        APPLICANT: Michel, Marie-Louise
;                   Mancine, Maryline
;        TITLE OF INVENTION: Nucleotide Vector, Composition
;                            Containing Such Vector, and Vaccine for Immunization
;                            Against Hepatitis
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;                        Dunner
;             STREET: 1300 I Street, N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: USA
;             ZIP: 20005-3315
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/142,358
;             FILING DATE: 10-May-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/799,569
;             FILING DATE: 12-FEB-1997
;             APPLICATION NUMBER: US 08/706,337
;             FILING DATE: 30-AUG-1996
;             APPLICATION NUMBER: US 08/633,821
;             FILING DATE: 22-APR-1996
;             APPLICATION NUMBER: FR 94/00483
;             FILING DATE: 27-APR-1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Meyers, Kenneth J
;             REGISTRATION NUMBER: 25,146
;             REFERENCE/DOCKET NUMBER: 03495.0128-01000
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (202) 408-4000
;             TELEFAX: (202) 408-4400
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5618 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-142-358-1

  Query Match             85.4%;  Score 110.2;  DB 9;  Length 5618;
```

```
  Best Local Similarity   93.5%;  Pred. No. 7.4e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    1900 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 1959

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    1960 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 2019

Qy     123 GCG 125
           |||
Db    2020 GCG 2022


RESULT 8
US-09-837-297-5
; Sequence 5, Application US/09837297
; Patent No. US20010049145A1
; GENERAL INFORMATION:
;  APPLICANT: RYU, WANG SHICK
;  TITLE OF INVENTION: Hepatitis B virus vectors for gene therapy
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/837,297
;  CURRENT FILING DATE:  2001-04-19
;  PRIOR APPLICATION NUMBER: KR2000-21070
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 5
;   LENGTH: 7991
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: R712: pCMV-HBV/GFP3.2 Full Sequence
US-09-837-297-5

  Query Match             85.4%;  Score 110.2;  DB 10;  Length 7991;
  Best Local Similarity   93.5%;  Pred. No. 8.1e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    2465 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 2524

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    2525 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 2584

Qy     123 GCG 125
           |||
Db    2585 GCG 2587


RESULT 9
US-09-837-297-3
; Sequence 3, Application US/09837297
; Patent No. US20010049145A1
; GENERAL INFORMATION:
;  APPLICANT: RYU, WANG SHICK
;  TITLE OF INVENTION: Hepatitis B virus vectors for gene therapy
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/837,297
;  CURRENT FILING DATE:  2001-04-19
;  PRIOR APPLICATION NUMBER: KR2000-21070
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 3
;   LENGTH: 8007
;   TYPE: DNA
;   ORGANISM: HBV
;   FEATURE:
;   NAME/KEY: gene
;   LOCATION: (1)..(8007)
;   OTHER INFORMATION: Prototype vector of HBV
US-09-837-297-3

  Query Match             85.4%;  Score 110.2;  DB 10;  Length 8007;
  Best Local Similarity   93.5%;  Pred. No. 8.1e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    2481 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 2540

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    2541 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 2600

Qy     123 GCG 125
           |||
Db    2601 GCG 2603


RESULT 10
US-09-837-297-4
; Sequence 4, Application US/09837297
; Patent No. US20010049145A1
; GENERAL INFORMATION:
;  APPLICANT: RYU, WANG SHICK
;  TITLE OF INVENTION: Hepatitis B virus vectors for gene therapy
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/837,297
;  CURRENT FILING DATE:  2001-04-19
;  PRIOR APPLICATION NUMBER: KR2000-21070
;  PRIOR FILING DATE: 2000-04-20
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: KopatentIn 1.71
; SEQ ID NO 4
;   LENGTH: 8717
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: R711: pCMV-HBV/GFP Full Sequence
US-09-837-297-4

  Query Match             85.4%;  Score 110.2;  DB 10;  Length 8717;
  Best Local Similarity   93.5%;  Pred. No. 8.3e-31;
  Matches  115;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  ||||||| ||| ||||||||||||||||||| |||||||||| ||||||||
Db    3191 CTTTCTGTGTAAACAATACCTGAACCTTTACCCCGTTGCCCGGCAACGGCCAGGTCTGTG 3250

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||||
Db    3251 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGTCATGGGCCATCAGCGCAT 3310

Qy     123 GCG 125
           |||
Db    3311 GCG 3313


RESULT 11
US-10-209-264-1
; Sequence 1, Application US/10209264
; Publication No. US20030003111A1
;   GENERAL INFORMATION:
;        APPLICANT: Oon, Chong Jin
;                   Lim, Gek Keow
;                   Zhao, Yi
;                   Chen, Wei Ning
;        TITLE OF INVENTION: A MUTANT HUMAN HEPATITIS B VIRAL STRAIN AND
;                            USES THEREOF
```

```
;        NUMBER OF SEQUENCES: 11
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Ladas & Parry
;             STREET: 26 West 61 Street
;             CITY: New York
;             STATE: New York
;             COUNTRY: USA
;             ZIP: 10023
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/209,264
;             FILING DATE: 31-Jul-2002
;             CLASSIFICATION: 435
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/SG98/00046
;             FILING DATE: 19-JAN-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mass, Clifford J.
;             REGISTRATION NUMBER: 30,086
;             REFERENCE/DOCKET NUMBER: U-013109-7
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (212) 708-1800
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3215 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: circular
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-209-264-1

  Query Match             81.4%;  Score 105;  DB 9;  Length 3215;
  Best Local Similarity   91.7%;  Pred. No. 5.4e-29;
  Matches  111;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  |||||||  ||  ||||||||||||||||||  |||||||| | |||||  ||
Db    1116 CTTTCTGTGTAAACAATATCTGAACCTTTACCCCGTTGCCCGGCAACGGTCCGGTCTCTG 1175

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1176 CCAAGTGTTTGCTGACGCAACCCCCACTGGATGGGGCTTGGCCATAGGCCATCAGCGCAT 1235

Qy     123 G 123
           |
Db    1236 G 1236


RESULT 12
US-09-897-511A-9
; Sequence 9, Application US/09897511A
; Publication No. US20030092882A1
; GENERAL INFORMATION:
;  APPLICANT: Bremel, Robert
;  APPLICANT:  Miller, Linda
;  APPLICANT:  Bleck, Gregory
;  TITLE OF INVENTION: Host Cells Containing Multiple Integrating Vectors
;  FILE REFERENCE: GALA-06416
;  CURRENT APPLICATION NUMBER: US/09/897,511A
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: 60/215,925
;  PRIOR FILING DATE: 2000-07-03
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 5130
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-897-511A-9

  Query Match             76.7%;  Score 99;  DB 9;  Length 5130;
  Best Local Similarity   87.8%;  Pred. No. 1e-26;
  Matches  108;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  |||||||  ||  ||||||||||||||||||  |||||||| | |||||  ||
Db    1995 CTTTCTGTGTAAACAATATCTGAACCTTTACCCCGTTGCCCGGCAACGGTCAGGTCTCTG 2054

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||| ||||||||||| || ||||||   |||||
Db    2055 CCAAGTGTTTGCTGACGCAACCCCCACTGGATGGGGCTTGGCTATCGGCCATAGCCGCAT 2114

Qy     123 GCG 125
           |||
Db    2115 GCG 2117


RESULT 13
US-09-897-006-9
; Sequence 9, Application US/09897006
; Patent No. US20020106729A1
; GENERAL INFORMATION:
;  APPLICANT: Bleck, Gregory
;  TITLE OF INVENTION: Expression Vectors
;  FILE REFERENCE: GALA-06415
;  CURRENT APPLICATION NUMBER: US/09/897,006
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: 60/215,851
;  PRIOR FILING DATE: 2000-07-03
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 5130
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic
US-09-897-006-9

  Query Match             76.7%;  Score 99;  DB 10;  Length 5130;
  Best Local Similarity   87.8%;  Pred. No. 1e-26;
  Matches  108;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||  |||||||  ||  ||||||||||||||||||  |||||||| | |||||  ||
Db    1995 CTTTCTGTGTAAACAATATCTGAACCTTTACCCCGTTGCCCGGCAACGGTCAGGTCTCTG 2054

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           |||||||||||||||||||||||||||||| ||||||||||| || ||||||   |||||
Db    2055 CCAAGTGTTTGCTGACGCAACCCCCACTGGATGGGGCTTGGCTATCGGCCATAGCCGCAT 2114

Qy     123 GCG 125
           |||
Db    2115 GCG 2117


RESULT 14
US-09-466-035-55
; Sequence 55, Application US/09466035
; Patent No. US20020165172A1
;  GENERAL INFORMATION:
;       APPLICANT: SALLBERG, MATTI
;                  MILICH, DAVID R.
;                  LEE, WILLIAM T.L.
;       TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
;                           INTRACELLULAR DISEASES
;       NUMBER OF SEQUENCES: 86
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Robins & Pasternak LLP
;            STREET: 545 Middlefield Road, Suite 180
```

```
;             CITY: Menlo Park
;             STATE: California
;             COUNTRY: U.S.
;             ZIP: 94025
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/466,035
;             FILING DATE: 17-Dec-1999
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Pasternak, Dahna S.
;             REGISTRATION NUMBER: 41,411
;             REFERENCE/DOCKET NUMBER: 2300-1231.01
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650-325-7812
;             TELEFAX: 650-325-7823
;             TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 55:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 67 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-466-035-55

  Query Match             40.6%;  Score 52.4;  DB 9;  Length 67;
  Best Local Similarity   98.1%;  Pred. No. 6.2e-10;
  Matches   53;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      72 TGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCG 125
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      14 TGCTGACGCAACCCCCACTGGCTGGGGCTTAGCCATAGGCCATCAGCGCATGCG 67


RESULT 15
US-09-912-679-55
; Sequence 55, Application US/09912679
; Patent No. US20020141974A1
;   GENERAL INFORMATION:
;        APPLICANT: Jolly, Douglas J.
;                   Chang, Stephen M.W.
;                   Lee, William T.L.
;                   Townsend, Kay
;                   O'Dea, Joanne
;        TITLE OF INVENTION: HEPATITIS THERAPEUTICS
;        NUMBER OF SEQUENCES: 84
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed and Berry
;             STREET: 6300 Columbia Center, 701 Fifth Avenue
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: U.S.
;             ZIP: 98104
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/912,679
;             FILING DATE: 07-Jun-1995
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMasters, David D.
;             REGISTRATION NUMBER: 33,963
;             REFERENCE/DOCKET NUMBER: 930049.407C5
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 206-622-4900
;             TELEFAX: 206-682-6031
;             TELEX: 3723836
;    INFORMATION FOR SEQ ID NO: 55:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 67 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-912-679-55

  Query Match             40.6%;  Score 52.4;  DB 10;  Length 67;
  Best Local Similarity   98.1%;  Pred. No. 6.2e-10;
  Matches   53;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      72 TGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCATGCG 125
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      14 TGCTGACGCAACCCCCACTGGCTGGGGCTTAGCCATAGGCCATCAGCGCATGCG 67


Search completed: May 21, 2003, 04:39:45
Job time : 122 secs
```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 21, 2003, 02:33:37 ; Search time 211 Seconds
(without alignments)
1376.814 Million cell updates/sec

Title: US-09-689-430-1_COPY_150_278
Perfect score: 129
Sequence: 1 ctctttctaagtaaacagta.....gccatcagcgcatgcggatc 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	129	100.0	7914	22	AAF84648	Plasmid DLZ7 encod
	2	129	100.0	7944	22	AAF84647	Plasmid DLZ6 encod
	3	123	95.3	587	18	AAT73163	Post-translational
	4	123	95.3	587	20	AAV69745	HBV post-transcrip
	5	123	95.3	909	24	AAH77169	Regulatory and cod
	6	123	95.3	3221	20	AAZ23281	DNA sequence of th
	7	123	95.3	4084	22	AAD14316	Hepatitis B virus
	8	123	95.3	4496	22	AAD14317	Hepatitis B virus
c	9	123	95.3	4525	20	AAV69746	Nucleotide sequenc
	10	123	95.3	6371	20	AAZ23285	DNA sequence of pl
	11	123	95.3	6371	20	AAZ23292	DNA sequence of pl
	12	123	95.3	6375	20	AAZ23294	DNA sequence of pl
	13	123	95.3	9325	20	AAZ23282	DNA sequence of pl
	14	123	95.3	9354	18	AAT73164	cDNA encoding huma
	15	123	95.3	9859	20	AAZ23286	DNA sequence of pl
	16	121.4	94.1	1355	4	AAN30059	Sequence encoding
	17	121.4	94.1	3183	4	AAN30035	DNA contg. surface
	18	121.4	94.1	3200	5	AAN40243	DNA encoding pre H
	19	121.4	94.1	3200	8	AAN70165	Entire nucleotide
	20	121.4	94.1	3221	22	AAH42374	Nucleotide sequenc
	21	119.8	92.9	3221	24	ABK23902	Bee venom phosphol
	22	119.8	92.9	7463	20	AAZ23293	DNA sequence of HB
	23	116.6	90.4	3215	21	AAZ29453	Mutant hepatitis B
	24	113.4	87.9	3220	21	AAZ88924	Hepatitis B virus
	25	111.8	86.7	2526	22	AAD09101	Hepatitis B virus
	26	111.8	86.7	3248	22	AAH77562	HBV genotype G str
	27	111.8	86.7	3248	22	AAD09091	Hepatitis B virus
	28	111.4	86.4	1500	20	AAV82697	Fulminant hepatiti
	29	111.2	86.2	1500	20	AAV82693	Fulminant hepatiti
	30	111.2	86.2	1500	20	AAV82694	Fulminant hepatiti
	31	110.2	85.4	306	24	ABK29876	Wild type hepatiti
c	32	110.2	85.4	1262	2	AAN10005	Sequence of Ava II
	33	110.2	85.4	1395	20	AAV82688	Fulminant hepatiti
	34	110.2	85.4	1500	20	AAV82686	Fulminant hepatiti
	35	110.2	85.4	1500	20	AAV82706	Wild type hepatiti
	36	110.2	85.4	2342	10	AAN93072	Sequence encoding
	37	110.2	85.4	3182	24	AAD31765	Hepatitis B virus
	38	110.2	85.4	5618	16	AAQ88310	Plasmid pRc/CMV-HB
	39	110.2	85.4	7991	24	AAS16094	HBV viral vector p
	40	110.2	85.4	8007	24	AAS16092	HBV viral vector p
	41	110.2	85.4	8717	24	AAS16093	HBV viral vector p
	42	108.6	84.2	1445	20	AAV82690	Fulminant hepatiti
	43	108.6	84.2	1445	20	AAV82692	Fulminant hepatiti
	44	108.6	84.2	1500	20	AAV82689	Fulminant hepatiti
	45	108.6	84.2	1500	20	AAV82695	Fulminant hepatiti

ALIGNMENTS

RESULT 1
AAF84648
ID AAF84648 standard; DNA; 7914 BP.
XX
AC AAF84648;
XX
DT 29-JUN-2001 (first entry)
XX
DE Plasmid DLZ7 encoding a canine B-domain deleted factor VIII.
XX
KW Adeno-associated virus vector; B-domain; factor VIII; haemophilia A;
KW coagulation disorder; ss.
XX
OS Synthetic.
OS Canis sp.
OS Hepatitis B virus.
XX
FH Key Location/Qualifiers
FT misc_feature 1..144
FT /*tag= a
FT /note= "inverted terminal repeat"
FT enhancer 149..278
FT /*tag= b
FT /note= "hepatitis B virus EnhI enhancer"
FT CDS 435..4730
FT /*tag= c
FT /note= "human B-domain deleted factor VIII"
FT polyA_signal 4804..4884
FT /*tag= d
FT /note= "TK polyA sequence"
FT misc_feature 4885..5048
FT /*tag= e

```
FT                  /note= "inverted terminal repeat"
XX
PN   WO200127303-A1.
XX
PD   19-APR-2001.
XX
PF   12-OCT-2000; 2000WO-US28221.
XX
PR   12-OCT-1999;   99US-0158780.
XX
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Walsh CE,  Chao H,  Burstein H,  Lynch CM,  Stepan AM,  Munson K;
XX
DR   WPI; 2001-273781/28.
DR   P-PSDB; AAB67960.
XX
PT   New recombinant adeno-associated virus vector, useful for treating
PT   haemophilia A, comprises heterologous nucleotide sequence encoding
PT   B-domain deleted human factor VIII operably linked with liver-preferred
PT   expression control element -
XX
PS   Disclosure; Fig 6; 87pp; English.
XX
CC   The specification describes a recombinant adeno-associated virus (rAAV)
CC   vector. The vector comprises a heterologous nucleotide sequence
CC   encoding B-domain deleted factor VIII operably linked with at least one
CC   enhancer and at least one promoter. The method results in the production
CC   of high titer rAAV vector stocks carrying the B-domain deleted factor
CC   VIII transgenes and expression cassettes, which generate adequate titers
CC   of virus for in vivo administration. The recombinant vectors are useful
CC   for treating haemophilia A, where the liver expresses the encoded
CC   B-domain deleted factor VIII, which is secreted into the blood. They are
CC   also useful for the treatment of other coagulation disorders. The
CC   present sequence encodes a B-domain deleted factor VIII.
XX
SQ   Sequence 7914 BP; 2055 A; 1994 C; 1950 G; 1915 T; 0 other;

  Query Match            100.0%;  Score 129;  DB 22;  Length 7914;
  Best Local Similarity  100.0%;  Pred. No. 9.2e-35;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 209

Qy      61 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 269

Qy     121 ATGCGGATC 129
           |||||||||
Db     270 ATGCGGATC 278


RESULT 2
AAF84647
ID   AAF84647 standard; DNA; 7944 BP.
XX
AC   AAF84647;
XX
DT   29-JUN-2001  (first entry)
XX
DE   Plasmid DLZ6 encoding human B-domain deleted factor VIII.
XX
KW   Adeno-associated virus vector; B-domain; factor VIII; haemophilia A;
KW   coagulation disorder; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Hepatitis B virus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..146
FT                   /*tag=  a
FT                   /note= "inverted terminal repeat"
FT   enhancer        150..278
FT                   /*tag=  b
FT                   /note= "hepatitis B virus EnhI enhancer"
FT   CDS             420..4835
FT                   /*tag=  c
FT                   /note= "human B-domain deleted factor VIII"
FT   polyA_signal    4840..4914
FT                   /*tag=  d
FT                   /note= "TK polyA sequence"
FT   misc_feature    4916..5084
FT                   /*tag=  e
FT                   /note= "inverted terminal repeat"
XX
PN   WO200127303-A1.
XX
PD   19-APR-2001.
XX
PF   12-OCT-2000; 2000WO-US28221.
XX
PR   12-OCT-1999;   99US-0158780.
XX
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Walsh CE,  Chao H,  Burstein H,  Lynch CM,  Stepan AM,  Munson K;
XX
DR   WPI; 2001-273781/28.
DR   P-PSDB; AAB67959.
XX
PT   New recombinant adeno-associated virus vector, useful for treating
PT   haemophilia A, comprises heterologous nucleotide sequence encoding
PT   B-domain deleted human factor VIII operably linked with liver-preferred
PT   expression control element -
XX
PS   Claim 64; Fig 1; 87pp; English.
XX
CC   The specification describes a recombinant adeno-associated virus (rAAV)
CC   vector. The vector comprises a heterologous nucleotide sequence
CC   encoding B-domain deleted factor VIII operably linked with at least one
CC   enhancer and at least one promoter. The method results in the production
CC   of high titer rAAV vector stocks carrying the B-domain deleted factor
CC   VIII transgenes and expression cassettes, which generate adequate titers
CC   of virus for in vivo administration. The recombinant vectors are useful
CC   for treating haemophilia A, where the liver expresses the encoded
CC   B-domain deleted factor VIII, which is secreted into the blood. They are
CC   also useful for the treatment of other coagulation disorders. The
CC   present sequence encodes a B-domain deleted factor VIII.
XX
SQ   Sequence 7944 BP; 2142 A; 1902 C; 1909 G; 1991 T; 0 other;

  Query Match            100.0%;  Score 129;  DB 22;  Length 7944;
  Best Local Similarity  100.0%;  Pred. No. 9.2e-35;
  Matches  129;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     150 CTCTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTG 209

Qy      61 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     210 TGCCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGC 269

Qy     121 ATGCGGATC 129
           |||||||||
Db     270 ATGCGGATC 278


RESULT 3
AAT73163
ID   AAT73163 standard; cDNA; 587 BP.
XX
```

```
AC   AAT73163;
XX
DT   08-APR-1998  (first entry)
XX
DE   Post-translational regulatory element (PRE) of the Hepatitis B virus.
XX
KW   Post-translational regulatory element; PRE; enhancer II; intronless gene;
KW   surface antigen gene; cytoplasmic accumulation; targeted delivery;
KW   near consensus splice sequence; blood coagulation factor; factor VIII;
KW   factor IX; ss.
XX
OS   Hepatitis B virus.
XX
PN   WO9733994-A1.
XX
PD   18-SEP-1997.
XX
PF   10-MAR-1997;   97WO-US03561.
XX
PR   11-MAR-1996;   96US-0683839.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Bidlingmaier S,  Ill CR;
XX
DR   WPI; 1997-470874/43.
XX
PT   Vector for increased expression of intronless genes - comprises
PT   intronless gene with at least one near consensus splice sequence, a
PT   promoter and at least one viral cis-acting post-transcriptional
PT   regulatory element
XX
PS   Claim 3; Page 21; 59pp; English.
XX
CC   The present sequence represents a post-translational regulatory element
CC   (PRE) of the Hepatitis B virus. This sequence encompasses enhancer II,
CC   and is within the transcribed portion of the surface antigen gene. This
CC   PRE sequence has been shown to function in cis to increase the
CC   steady-state levels of surface gene transcripts by facilitating
CC   cytoplasmic accumulation of these transcripts. The present PRE sequence
CC   was used to create a novel vector, comprising an intronless gene
CC   containing  or more near consensus splice sequences operably linked to a
CC   promoter sequence so that the gene is transcribed in a cell. One or more
CC   copies of a viral cis-acting PRE are also cloned into the vector, and are
CC   transcribed along with the gene, causing export of the gene transcript
CC   from the nucleus into the cytoplasm of the cell. The vector can be used
CC   to increase the expression of an intronless gene containing at least one
CC   near consensus splice sites, preferably cDNA encoding a blood coagulation
CC   factor, particularly Factor VIII or IX. The complex allows the targeted
CC   delivery of the vector to a specific cell, e.g. hepatocytes when the
CC   ligand is an asialoglycoprotein which binds the asialoglycoprotein
CC   receptor present on their surface.
XX
SQ   Sequence 587 BP; 97 A; 199 C; 145 G; 146 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 18;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 61

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 121

Qy     123 GCG 125
           |||
Db     122 GCG 124


RESULT 4
AAV69745
ID   AAV69745 standard; cDNA; 587 BP.
XX
AC   AAV69745;
XX
DT   04-FEB-1999  (first entry)
XX
DE   HBV post-transcriptional regulatory element (PRE) sequence.
XX
KW   Hepatitis B virus; post-transcriptional regulatory element; PRE; HBV;
KW   viral transcript; ss.
XX
OS   Hepatitis b virus.
XX
PN   US5843770-A.
XX
PD   01-DEC-1998.
XX
PF   11-MAR-1996;   96US-0613861.
XX
PR   11-MAR-1996;   96US-0613861.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Gonzales JEN,  Ill CR;
XX
DR   WPI; 1999-044589/04.
XX
PT   Hepatitis B virus antisense vector  - directed against cis-acting
PT   post-transcriptional regulatory element
XX
PS   Claim 1; Columns 13-14; 12pp; English.
XX
CC   This sequence represents a hepatitis B virus cis-acting post-
CC   transcriptional regulatory element (PRE). The invention provides a vector
CC   encoding one or more antisense transcripts that are complementary to all
CC   or part of a HBV PRE where the PRE directs export of viral transcripts
CC   from the nucleus to the cytoplasm of a cell.  A molecular complex
CC   comprising the above vector can be releasably linked to a conjugate of
CC   a nucleic acid binding agent and a ligand that binds to a component on
CC   the surface of a cell. The vector can be delivered to cells in vitro or
CC   in vivo to inhibit production of viruses having PRE sequences.
XX
SQ   Sequence 587 BP; 97 A; 199 C; 145 G; 146 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 61

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 121

Qy     123 GCG 125
           |||
Db     122 GCG 124


RESULT 5
AAH77169
ID   AAH77169 standard; DNA; 909 BP.
XX
AC   AAH77169;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Regulatory and coding region of the X15 component in the X-myc construct.
XX
KW   Transgenic mouse; cancer; oncogene; bicistronic hepatitis B virus; HBV;
```

```
KW   X15-c-myc transgene; hepatocellular carcinoma; malignant liver tumour;
KW   X15; c-myc; murine; HBX; carcinogen; ds.
XX
OS   Hepatitis B virus.
XX
PN   US6274788-B1.
XX
PD   14-AUG-2001.
XX
PF   02-FEB-1999;   99US-0243282.
XX
PR   23-SEP-1998;   98IN-0002858.
XX
PA   (ITGE-) INT CENT GENETIC ENG & BIOTECHNOLOGY.
PA   (NAIM-) NAT INST IMMUNOLOGY.
XX
PI   Kumar V,  Singh M,  Totey S,  Anand R;
XX
DR   WPI; 2002-009266/01.
XX
PT   New bicistronic hepatitis B virus (HBV) X15-c-myc transgene, useful for
PT   producing transgenic mouse model systems for human hepatocellular
PT   carcinoma, comprises HBV X15 transgene and c-myc transgene -
XX
PS   Claim 3; Fig 3; 12pp; English.
XX
CC   This polynucleotide represents the sequence of the regulatory and coding
CC   regions of the X15 component in the X-myc construct. The invention
CC   relates to a bicistronic hepatitis B virus (HBV) X15-c-myc transgene,
CC   comprising of the HBV X15 gene and c-myc gene. The myc gene is known to
CC   be an activatable oncogene. The transgene encodes a truncated HBV X15
CC   protein that has amino acids 58-154 of HBV X15 and a murine c-myc
CC   protein, respectively. A transgenic mouse containing the transgene
CC   construct is useful for screening a candidate substance (CS), to
CC   determine whether CS promotes hepatocellular carcinoma. This is
CC   determined by exposing a transgenic mouse to CS, and monitoring the mouse
CC   for the development of hepatocellular carcinoma, where an increase in the
CC   development of hepatocellular carcinoma in the transgenic mouse exposed
CC   to CS compared to the development of hepatocellular carcinoma in a
CC   transgenic mouse not exposed to CS, indicates that CS promotes
CC   hepatocellular carcinoma. The transgenic mice can be employed as a source
CC   for cell and tissue culture. The transgenic animal models comprising of
CC   the HBV X15-c-myc transgene for hepatocellular carcinoma are superior to
CC   any transgenic animal model system for hepatocellular carcinoma in that
CC   the transgenic mice develop more aggressive and accelerated onset of
CC   malignant liver tumours in all lobes causing death of the affected
CC   animals in 20-22 weeks, that is faster than the time taken by the other
CC   transgenic animals to even develop a tumour.
XX
SQ   Sequence 909 BP; 210 A; 236 C; 211 G; 252 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 24;  Length 909;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     286 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 345

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     346 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 405

Qy     123 GCG 125
           |||
Db     406 GCG 408


RESULT 6
AAZ23281
ID   AAZ23281 standard; DNA; 3221 BP.
XX
AC   AAZ23281;
XX
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of the genome of HBV adw 2.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Disclosure; Columns 35-39; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
CC   The present sequence represents the DNA sequence of the genome of HBV
CC   adw 2 which comprises the pol gene sequences, X gene sequences and
CC   surface antigen gene (preS1/preS2/S gene) sequences.
XX
SQ   Sequence 3221 BP; 740 A; 869 C; 708 G; 904 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 3221;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1118 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1177

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1178 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1237

Qy     123 GCG 125
           |||
```

```
Db    1238 GCG 1240


RESULT 7
AAD14316
ID   AAD14316 standard; DNA; 4084 BP.
XX
AC   AAD14316;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Hepatitis B virus (HBV) 1.28 genome.
XX
KW   Hepatitis B virus; HBV; altered sensitivity; agent; detection;
KW   PCR primer; Hepatitis B surface antigen; HBsAg; ds.
XX
OS   Hepatitis B virus.
XX
PN   WO200157244-A1.
XX
PD   09-AUG-2001.
XX
PF   02-FEB-2001; 2001WO-AU00098.
XX
PR   03-FEB-2000; 2000US-0179948.
XX
PA   (MELB-) MELBOURNE HEALTH.
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Delaney W,  Locarnini SA,  Chen RYM,  Bartholomeusz A,  Isom H;
XX
DR   WPI; 2001-496926/54.
XX
PT   Detecting hepatitis B virus variant with altered sensitivity to agent,
PT   comprises infecting genetic construct containing replication competent
PT   genome to cells, contacting cells with agent and detecting replication
PT   of variant  -
XX
PS   Example 7; Fig 5A; 110pp; English.
XX
CC   The invention relates to a method of detecting variant hepatitis B virus
CC   (HBV) which exhibits altered sensitivity to agents. The method involves
CC   infecting a genetic construct containing a replication competent amount
CC   of the genome from variant HBV contained in or fused to a baculovirus
CC   genome; contacting cells with the agent to be tested; culturing cells
CC   under conditions sufficient for the variant HBV to replicate, express
CC   genetic sequences, and/or assemble, and/or release viral particles; and
CC   determining replication of variant HBV using viral-component-detection
CC   means. The method is useful for detecting variant HBV which exhibits
CC   altered sensitivity to agents. The present sequence is the HBV 1.28
CC   genome.
XX
SQ   Sequence 4084 BP; 920 A; 1105 C; 914 G; 1145 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 22;  Length 4084;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2886 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 2945

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2946 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 3005

Qy     123 GCG 125
           |||
Db    3006 GCG 3008


RESULT 8
AAD14317
ID   AAD14317 standard; DNA; 4496 BP.
XX
AC   AAD14317;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Hepatitis B virus (HBV) 1.5 genome.
XX
KW   Hepatitis B virus; HBV; altered sensitivity; agent; detection;
KW   PCR primer; Hepatitis B surface antigen; HBsAg; ds.
XX
OS   Hepatitis B virus.
XX
PN   WO200157244-A1.
XX
PD   09-AUG-2001.
XX
PF   02-FEB-2001; 2001WO-AU00098.
XX
PR   03-FEB-2000; 2000US-0179948.
XX
PA   (MELB-) MELBOURNE HEALTH.
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Delaney W,  Locarnini SA,  Chen RYM,  Bartholomeusz A,  Isom H;
XX
DR   WPI; 2001-496926/54.
XX
PT   Detecting hepatitis B virus variant with altered sensitivity to agent,
PT   comprises infecting genetic construct containing replication competent
PT   genome to cells, contacting cells with agent and detecting replication
PT   of variant  -
XX
PS   Example 7; Fig 5B; 110pp; English.
XX
CC   The invention relates to a method of detecting variant hepatitis B virus
CC   (HBV) which exhibits altered sensitivity to agents. The method involves
CC   infecting a genetic construct containing a replication competent amount
CC   of the genome from variant HBV contained in or fused to a baculovirus
CC   genome; contacting cells with the agent to be tested; culturing cells
CC   under conditions sufficient for the variant HBV to replicate, express
CC   genetic sequences, and/or assemble, and/or release viral particles; and
CC   determining replication of variant HBV using viral-component-detection
CC   means. The method is useful for detecting variant HBV which exhibits
CC   altered sensitivity to agents. The present sequence is the HBV 1.5
CC   genome.
XX
SQ   Sequence 4496 BP; 999 A; 1229 C; 1011 G; 1257 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 22;  Length 4496;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 136

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 196

Qy     123 GCG 125
           |||
Db     197 GCG 199


RESULT 9
AAV69746/c
ID   AAV69746 standard; cDNA; 4525 BP.
XX
AC   AAV69746;
XX
DT   04-FEB-1999  (first entry)
```

```
XX
DE   Nucleotide sequence of one copy antisense plasmid pCMVasPRE-RZ.
XX
KW   Hepatitis B virus; post-transcriptional regulatory element; PRE; HBV;
KW   viral transcript; pCMVasPRE-RZ; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5843770-A.
XX
PD   01-DEC-1998.
XX
PF   11-MAR-1996;   96US-0613861.
XX
PR   11-MAR-1996;   96US-0613861.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Gonzales JEN,  Ill CR;
XX
DR   WPI; 1999-044589/04.
XX
PT   Hepatitis B virus antisense vector  - directed against cis-acting
PT   post-transcriptional regulatory element
XX
PS   Claim 3; Columns 13-18; 12pp; English.
XX
CC   This represents the nucleotide sequence of an one copy antisense plasmid
CC   pCMVasPRE-RZ. The plasmid contains a hepatitis B virus cis-acting post-
CC   transcriptional regulatory element (PRE). The invention provides such a
CC   vector encoding one or more antisense transcripts that are complementary
CC   to all or part of a HBV PRE where the PRE directs export of viral
CC   transcripts from the nucleus to the cytoplasm of a cell.  A molecular
CC   complex comprising the above vector can be releasably linked to a
CC   conjugate of a nucleic acid binding agent and a ligand that binds to a
CC   component on the surface of a cell. The vector can be delivered to cells
CC   in vitro or in vivo to inhibit production of viruses having PRE
CC   sequences.
XX
SQ   Sequence 4525 BP; 1115 A; 1132 C; 1168 G; 1110 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 4525;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-33;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3727 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 3668

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3667 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 3608

Qy     123 GCG 125
           |||
Db    3607 GCG 3605


RESULT 10
AAZ23285
ID   AAZ23285 standard; DNA; 6371 BP.
XX
AC   AAZ23285;
XX
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of plasmid pTHBVT.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Example 1; Columns 47-54; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
CC   The present sequence represents the DNA sequence of the plasmid pTHBVT
CC   which comprises HBV sequences.
XX
SQ   Sequence 6371 BP; 1568 A; 1650 C; 1485 G; 1668 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 6371;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1423

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1483

Qy     123 GCG 125
           |||
Db    1484 GCG 1486


RESULT 11
AAZ23292
ID   AAZ23292 standard; DNA; 6371 BP.
XX
AC   AAZ23292;
XX
```

```
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of plasmid pTHBVTX-.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Example 2; Columns 63-70; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
CC   The present sequence represents the DNA sequence of the plasmid pTHBVTX-.
XX
SQ   Sequence 6371 BP; 1567 A; 1649 C; 1485 G; 1670 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 6371;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1423

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1483

Qy     123 GCG 125
          |||
Db    1484 GCG 1486
```

```
RESULT 12
AAZ23294
ID   AAZ23294 standard; DNA; 6375 BP.
XX
AC   AAZ23294;
XX
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of plasmid pTHBVTP-.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Example 3; Columns 75-82; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
CC   The present sequence represents the DNA sequence of the plasmid pTHBVTP-.
XX
SQ   Sequence 6375 BP; 1568 A; 1652 C; 1487 G; 1668 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 6375;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1423

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1483

Qy     123 GCG 125
           |||
Db    1484 GCG 1486


RESULT 13
AAZ23282
ID   AAZ23282 standard; DNA; 9325 BP.
XX
AC   AAZ23282;
XX
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of plasmid pTHBV-d.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Example 1; Columns 39-48; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
CC   The present sequence represents the DNA sequence of the plasmid pTHBV-d
CC   which comprises HBV sequences.
XX
SQ   Sequence 9325 BP; 2227 A; 2448 C; 2132 G; 2518 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 9325;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1423

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1483

Qy     123 GCG 125
           |||
Db    1484 GCG 1486


RESULT 14
AAT73164
ID   AAT73164 standard; cDNA; 9354 BP.
XX
AC   AAT73164;
XX
DT   08-APR-1998  (first entry)
XX
DE   cDNA encoding human B-domain deleted factor VIII.
XX
KW   Post-translational regulatory element; PRE; enhancer II; intronless gene;
KW   surface antigen gene; cytoplasmic accumulation; targeted delivery;
KW   near consensus splice sequence; blood coagulation factor; factor VIII;
KW   factor IX; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2965..7380
FT                   /*tag= a
FT   misc_feature    5165..5174
FT                   /*tag= b
FT                   /note= "5' near consensus site"'
FT   misc_feature    5695..5703
FT                   /*tag= c
FT                   /note= "5' near consensus site"
FT   misc_feature    6320..6328
FT                   /*tag= d
FT                   /note= "5' near consensus site"
FT   misc_feature    6595..5603
FT                   /*tag= e
FT                   /note= "5' near consensus site"
FT   misc_feature    7045..7053
FT                   /*tag= f
FT                   /note= "5' near consensus site"
FT   misc_feature    7143..7152
FT                   /*tag= g
FT                   /note= "5' near consensus site"
FT   misc_feature    3296..3312
FT                   /*tag= h
FT                   /note= "3' near consensus site"
FT   misc_feature    4798..4817
FT                   /*tag= i
FT                   /note= "3' near consensus site"
FT   misc_feature    5023..5045
FT                   /*tag= j
FT                   /note= "3' near consensus site"
FT   misc_feature    5333..5355
FT                   /*tag= k
FT                   /note= "3' near consensus site"
FT   misc_feature    5520..5538
FT                   /*tag= l
FT                   /note= "3' near consensus site"
FT   misc_feature    5604..5632
FT                   /*tag= m
FT                   /note= "3' near consensus site"
FT   misc_feature    5717..5745
```

```
FT                   /*tag=  n
FT                   /note= "3' near consensus site"
FT   misc_feature    6239..6258
FT                   /*tag=  o
FT                   /note= "3' near consensus site"
FT   misc_feature    6658..6682
FT                   /*tag=  p
FT                   /note= "3' near consensus site"
FT   misc_feature    7159..7176
FT                   /*tag=  q
FT                   /note= "3' near consensus site"
FT   misc_feature    7196..7209
FT                   /*tag=  r
FT                   /note= "3' near consensus site"
FT   misc_feature    7289..7315
FT                   /*tag=  s
FT                   /note= "3' near consensus site"
FT   misc_feature    7411..7429
FT                   /*tag=  t
FT                   /note= "3' near consensus site"
FT   misc_feature    7611..8197
FT                   /*tag=  u
FT                   /note= "PRE sequence"
XX
PN   WO9733994-A1.
XX
PD   18-SEP-1997.
XX
PF   10-MAR-1997;   97WO-US03561.
XX
PR   11-MAR-1996;   96US-0683839.
XX
PA   (IMMU-) IMMUNE RESPONSE CORP.
XX
PI   Bidlingmaier S,  Ill CR;
XX
DR   WPI; 1997-470874/43.
DR   P-PSDB; AAW23414.
XX
PT   Vector for increased expression of intronless genes - comprises
PT   intronless gene with at least one near consensus splice sequence, a
PT   promoter and at least one viral cis-acting post-transcriptional
PT   regulatory element
XX
PS   Example 1; Pages 21-31; 59pp; English.
XX
CC   The present sequence represents human B-domain deleted factor VIII
CC   cDNA, and a post-translational regulatory element (PRE) of the
CC   Hepatitis B virus, which is present 3' of the STOP codon for factor VIII.
CC   PRE sequences have been shown to function in cis to increase the
CC   steady-state levels of surface gene transcripts by facilitating
CC   cytoplasmic accumulation of these transcripts. The present sequence
CC   is part of a novel vector, comprising an intronless gene containing
CC   1 or more near consensus splice sequences operably linked to a
CC   promoter sequence so that the gene is transcribed in a cell.
CC   Intronless gene transcripts which contain near consensus splice site
CC   sequences are believed to get tied up in the nucleus of the cell where
CC   splicing occurs, rather than being transported to the cytoplasm where
CC   they can be translated into proteins. The PRE sequences are transcribed
CC   along with the gene, causing export of the gene transcript from the
CC   nucleus into the cytoplasm of the cell. The vector can be used
CC   to increase the expression of an intronless gene containing at least one
CC   near consensus splice sites, preferably cDNA encoding a blood coagulation
CC   factor, particularly Factor VIII or IX. The complex allows the targeted
CC   delivery of the vector to a specific cell, e.g. hepatocytes when the
CC   ligand is an asialoglycoprotein which binds the asialoglycoprotein
CC   receptor present on their surface.
XX
SQ   Sequence 9354 BP; 2506 A; 2239 C; 2161 G; 2448 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 18;  Length 9354;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7612 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 7671

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7672 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 7731

Qy     123 GCG 125
          |||
Db    7732 GCG 7734


RESULT 15
AAZ23286
ID   AAZ23286 standard; DNA; 9859 BP.
XX
AC   AAZ23286;
XX
DT   31-JAN-2000  (first entry)
XX
DE   DNA sequence of plasmid pTHBVT-d.
XX
KW   Hepatitis B virus; HBV; recombinant; pol gene; X gene; surface antigen;
KW   liver; anti-viral; anti-tumor; gene therapy; single-gene defect;
KW   genetic disorder; familial hypercholesterolemia; neoplastic gene;
KW   ornithine transcarbamylase deficiency; ss.
XX
OS   Synthetic.
OS   Hepatitis b virus.
XX
PN   US5981274-A.
XX
PD   09-NOV-1999.
XX
PF   18-SEP-1996;   96US-0715808.
XX
PR   18-SEP-1996;   96US-0715808.
XX
PA   (CHAI/) CHAISOMCHIT S.
PA   (CHAN/) CHANG L.
PA   (TYRR/) TYRRELL D L J.
XX
PI   Chang L,  Chaisomchit S,  Tyrrell DLJ;
XX
DR   WPI; 1999-633330/54.
XX
PT   Recombinant hepatitis B virus genome containing heterologous gene
PT   sequences useful for treating liver infections -
XX
PS   Example 1; Columns 53-62; 53pp; English.
XX
CC   The invention relates to a recombinant hepatitis B virus genome (HBV)
CC   that comprises heterologous gene sequences which express at least one
CC   functional heterologous gene product. A host cell transfected with a
CC   recombinant HBV genome comprising pol gene sequences, X gene sequences
CC   and surface antigen gene (preS1/preS2/S gene) sequences and heterologous
CC   gene sequences can be used to express at least one functional
CC   heterologous gene product. The invention also provides a method for
CC   encapsidating a recombinant HBV genome. The recombinant HBV genomes are
CC   useful for the expression of functional heterologous gene products in
CC   liver cells. The vectors can be used for anti-viral, anti-tumor and/or
CC   gene therapy and particularly for the correction of inherited single-gene
CC   defects. Human genetic disorders which can be treated by expression of
CC   missing or mutant genes in the liver are familial hypercholesterolemia
CC   and ornithine transcarbamylase deficiency. Primary tumors of the liver
CC   may benefit from the expression of anti-neoplastic genes in the liver.
CC   Existing retroviral vectors and other animal viruses which are used to
CC   deliver foreign genes are not liver-specific with regard to their
CC   infection or expression unlike hepatitis B viral vectors. Human hepatitis
CC   B virus can be delivered through the circulation so there is no
CC   requirement for tissue culture for ex vivo liver-directed gene therapy.
```

```
CC   The present sequence represents the DNA sequence of the plasmid pTHBTV-d.
XX
SQ   Sequence 9859 BP; 2389 A; 2590 C; 2254 G; 2626 T; 0 other;

  Query Match             95.3%;  Score 123;  DB 20;  Length 9859;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-32;
  Matches  123;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CTTTCTAAGTAAACAGTACATGAACCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTG 1423

Qy      63 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CCAAGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCTTGGCCATAGGCCATCAGCGCAT 1483

Qy     123 GCG 125
          |||
Db    1484 GCG 1486


Search completed: May 21, 2003, 03:56:18
Job time : 223 secs
```